

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2002, 03:52:00 ; Search time 2120.21 Seconds
(without alignments)
60.819 Million cell updates/sec

Title: US-09-700-187-1
Perfect score: 12
Sequence: 1 ggattttacagt 12

Indexing table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: em_estfun: *
2: em_esthum: *
3: em_estin: *
4: em_estom: *
5: em_estpl: *
6: em_estba: *
7: em_estro: *
8: em_estov: *
9: em_hic: *
10: gb_est1: *
11: gb_est2: *
12: gb_hic: *
13: gb_gss: *
14: em_gss_fun: *
15: em_gss_hum: *
16: em_gss_inv: *
17: em_gss_pln: *
18: em_gss_pro: *
19: em_gss_rod: *
20: em_gss_vrt: *
21: em_gss_other: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	12	100.0	88	13	AZ925237 4910.ez32
C 2	12	100.0	118	10	AW217731 EST296445
C 3	12	100.0	126	13	AZ289196 RPT1-23-5
C 4	12	100.0	130	10	AI610647 tp20d02.x
C 5	12	100.0	147	10	BE145737 IL5-HF020
C 6	12	100.0	147	11	D25785 HUMGS04153
C 7	12	100.0	150	10	AA579315 nf36e06.s
C 8	12	100.0	152	10	AU180504 AU180504
C 9	12	100.0	155	13	AQ009485 CMT-HSP-2
C 10	12	100.0	165	10	AV333933 AV333933
C 11	12	100.0	165	11	BF881282 QV1-EF018
C 12	12	100.0	171	11	BF909645 PM3-UT005

c 13 12 100.0 172 10 AT029683
c 14 12 100.0 173 13 AZ818519
c 15 12 100.0 177 13 AQ985182
c 16 12 100.0 180 11 BF235518
c 17 12 100.0 184 10 AV054690
c 18 12 100.0 185 10 AV071084
c 19 12 100.0 185 10 BG357868
c 20 12 100.0 185 11 BG942699
c 21 12 100.0 186 10 AI345298
c 22 12 100.0 189 11 F29452
c 23 12 100.0 190 10 AV288772
c 24 12 100.0 193 10 AV285178
c 25 12 100.0 194 10 AI136924
c 26 12 100.0 202 11 D60959
c 27 12 100.0 207 10 AI572468
c 28 12 100.0 210 10 AV336005
c 29 12 100.0 216 13 B25087
c 30 12 100.0 217 10 AI563911
c 31 12 100.0 220 13 AQ278868
c 32 12 100.0 223 10 BE236288
c 33 12 100.0 223 10 BE236293
c 34 12 100.0 224 10 AV252038
c 35 12 100.0 224 10 BB512144
c 36 12 100.0 225 10 BB256670
c 37 12 100.0 226 11 BI220740
c 38 12 100.0 227 10 AA129692
c 39 12 100.0 227 11 BI293223
c 40 12 100.0 228 10 BB171532
c 41 12 100.0 228 10 BB310621
c 42 12 100.0 230 10 BB152264
c 43 12 100.0 231 10 AV368688
c 44 12 100.0 231 10 BB503511
c 45 12 100.0 231 10 BB510596

ALIGNMENTS

RESULT 1
AZ925237/c

LOCUS 88 bp DNA GSS 01-APR-2001
DEFINITION 4910.ez32kl8.sl Saccharomyces paradoxus N17 Saccharomyces paradoxus genomic clone 4910.ez32kl8.sl, DNA sequence.

ACCESSION AZ925237

VERSION AZ925237.1 GI:13496136

KEYWORDS GSS

SOURCE Saccharomyces paradoxus.

ORGANISM Saccharomyces paradoxus

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

REFERENCE 1 (bases 1 to 88)

AUTHORS Clifton,P.F., Hillier,L.W., Fulton,L., Graves,T., Miner,T., Gish

,W.R., Waterston,R.H. and Johnston,M.

TITLE Surveying Saccharomyces genomes to identify functional elements by

comparative DNA sequence analysis

JOURNAL Unpublished (2001)

COMMENT Contact: Johnston M

Department of Genetics

Washington University Medical School

Box 8232, 4566 Scott Ave., St. Louis, MO 63110, USA

Tel: 314 362 2735

Fax: 314 362 7855

Email: mjgenetics.wustl.edu

Class: random plasmid subclone.

Location/Qualifiers

1. .88

source

/organism="Saccharomyces paradoxus"

/strain="N17"

/db_xref="taxon:27291"

/clone="4910.ez32kl8.sl"

/clone_lib="Saccharomyces paradoxus N17"

/note="Random genomic sequence "

BASE COUNT 26 a 12 c 8 g 4 t

ORIGIN

Query Match 100.0%; Score 12; DB 13; Length 88;
 Best Local Similarity 100.0%; Pred. No. 7.3e+03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
 |||||
 Db 40 GGATTTTACAGT 29

RESULT 2

AW217731/c
 LOCUS AW217731 118 bp mRNA EST 18-MAY-2001
 DEFINITION EST296445 tomato flower buds 8 mm to pre-anthesis, Cornell University Lycopersicon esculentum cDNA clone cT0C6E5, mRNA sequence.

ACCESSION AW217731.1 GI:6528605

VERSION AW217731
 KEYWORDS EST.
 SOURCE tomato.

ORGANISM

Lycopersicon esculentum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
 Lycopersicon.

REFERENCE 1 (bases 1 to 118)

van der Hoeven, R.S., Bezzerides, J.L., Matern, A.L., Holt, I.E., Liang, F., Upton, J., Hansen, T., Craven, M.B., Bowman, C.L., Ahn, S., Ronning, C.M., Fraser, C.M., Martin, G.B., Giovannoni, J.J., and Tanksley, S.D.
 Generation of ESTs from tomato flower tissue
 Unpublished (1999)
 Contact: CUGI

-TITLE
JOURNAL
COMMENT

Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Email: <http://www.genome.clemson.edu/orders/index.html>
 5 prime sequence.

FEATURES

source

1. .118
 Location/Qualifiers
 /organism="Lycopersicon esculentum"
 /cultivar="TA496"
 /db_xref="taxon:4081"
 /clone_lib="cT0C6E5"
 /clone_lib="tomato flower buds 8 mm to pre-anthesis,
 Cornell University"
 /tissue_type="flower"

/dev_stage="buds 8mm-to-preanthesis"
 /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
 XhoI; supplier: Tanksley; Flower buds and flowers were
 taken from greenhouse plants (4-8 wks old, TA496). They
 were immediately frozen in liquid nitrogen and then
 size-separated while remaining frozen."

BASE COUNT
ORIGIN

49 a 21 c 16 g 32 t

Query Match 100.0%; Score 12; DB 10; Length 118;
 Best Local Similarity 100.0%; Pred. No. 7.2e+03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
 |||||
 Db 47 GGATTTTACAGT 36

RESULT 3

AZ289196
 LOCUS AZ289196 126 bp DNA GSS 27-JUL-2000
 DEFINITION RPCI-23-59B23.TJB RPCI-23 Mus musculus genomic clone RPCI-23-59B23,
 DNA sequence.
 ACCESSION AZ289196

VERSION

KEYWORDS

SOURCE

ORGANISM

AZ289196.1 GI:9530982

GSS.

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 126)

REFERENCE

AUTHORS

Zhao, S., Nierman, W., Feidblyum, T., Malek, J., Shatsman, S., Akinret,
 B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.
 and Fraser, C.M.

Mouse BAC End Sequences from Library RPCI-23

Unpublished (1999)

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC
 library availability, please contact Pieter de Jong
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.html>)
 or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html

Seq primer: SP6

Plate: 59 row: B column: 23

Class: BAC ends.

Location/Qualifiers

1. .126

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-23-59B23"

/clone_lib="RPCI-23"

/sex="Female"

/lab_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site_1:
 EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
 brain genomic DNA was isolated and partially digested
 with a combination of EcoRI and EcoRI Methylase. Size
 selected DNA was cloned into the pBACE3.6 vector at the
 EcoRI sites. The ligation products were transformed into
 DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT

ORIGIN

38 a 20 c 30 g 38 t

Query Match

100.0%; Score 12; DB 13; Length 126;

Best Local Similarity 100.0%; Pred. No. 7.1e+03;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12

|||||

Db 93 GGATTTTACAGT 104

RESULT 4

AI610647/c

LOCUS

DEFINITION

AI610647 130 bp mRNA EST 21-APR-1999
 tp20d02.x1 NC1_CGAP_Gas4 Homo sapiens cDNA clone IMAGE:2188323 3'
 similar to SW:ULC6_HCMVA P16836 HYPOTHETICAL PROTEIN UL126.
 ; contains element LTR5 repetitive element ; mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

AI610647.1 GI:4619814

EST.

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 130)

REFERENCE

AUTHORS

TITLE

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

JOURNAL
COMMENT

Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cqbbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality

Seq primer: -40up from Gibco

High quality sequence stop: 1.

FEATURES
source

Location/Qualifiers
1. .130
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2188323"
/clone_lib="NCL_CGAP_Gas4"
/tissue_type="poorly differentiated adenocarcinoma with
signet ring cell features"
/lab_host="DH10B"
/note="Organ: stomach; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.69 kb. Life Technologies catalog #:
11549-011"

BASE COUNT 36 a 21 c 28 g 45 t
ORIGIN

Query Match 100.0%; Score 12; DB 10; Length 130;
Best Local Similarity 100.0%; Pred. No. 7.1e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggattttacagt 12
|||||
Db 115 GGATTTTACAGT 104

RESULT 5

LOCUS BE145737/c 147 bp mRNA EST 21-JUN-2000
DEFINITION IL5-HT0207-231099-006-A01 HT0207 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE145737
VERSION BE145737.1 GI:8608461
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 147)

REFERENCE
AUTHORS Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.P., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE
COMMENT

20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=tl5-HT0207-231
099-006-A01&tl3=1999-10-23&tl4=1)
Seq primer: puc 18 forward
High quality sequence stop: 147.
High quality sequence stop: 147.

FEATURES
source

Location/Qualifiers
1. .147
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0207"
/dev_stage="Adult"
/note="Organ: head neck; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT 31 a 35 c 12 g 69 t
ORIGIN

Query Match 100.0%; Score 12; DB 10; Length 147;
Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggattttacagt 12
|||||
Db 132 GGATTTTACAGT 121

RESULT 6

LOCUS D25785/c 147 bp mRNA EST 30-NOV-1995
DEFINITION HUMGS04153 Human colon mucosa Homo sapiens cDNA clone cm1984 3',
mRNA sequence.

ACCESSION D25785
VERSION D25785.1 GI:500469
KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 147)

AUTHORS Okubo, K., Yoshii, J., Yokouchi, H., Kameyama, M. and Matsubara, K.
TITLE Global analysis of gene expression in colon mucosa: a large scale
random cDNA sequencing analysis

JOURNAL Unpublished (1994)

COMMENT Contact: Okubo, K., Itoh, K., Yoshii, J., Yokouchi, H. and Matsubara, K.
Institute for Molecular and Cellular Biology
Osaka University
3-1 Yamada-oka, Suita, Osaka 565, Japan.

FEATURES
source

Location/Qualifiers
1. .147
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="cm1984"
/clone_lib="Human colon mucosa"
/note="Adult male, tissue_type = colon mucosa "

BASE COUNT 63 a 24 c 34 g 26 t
ORIGIN

Query Match 100.0%; Score 12; DB 11; Length 147;
Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggattttacagt 12
|||||
Db 57 GGATTTTACAGT 46

RESULT 7
 AA579315 150 bp mRNA 12-SEP-1997
 LOCUS nf36e06.s1 NCI_CGAP_Pr2 Homo sapiens cDNA clone IMAGE:915874, mRNA
 DEFINITION sequence.
 ACCESSION AA579315
 VERSION AA579315.1 GI:2357499
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 150)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui, M.D.,
 Michael Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: David B. Krizman, Ph.D.
 DNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
 Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
 Insert Length: 578 Std Error: 0.00
 Seq primer: -40ml3 fwd. ET from Amersham
 High quality sequence stop: 126.

FEATURES

Location/Qualifiers
 1..150
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:915874"
 /clone_lib="NCI_CGAP_Pr2"
 /sex="Male"
 /dev_stage="45 years old"
 /lab_host="DH10B"
 /note="Vector: pAMP10; Site 1: NotI; Site 2: EcoRI; 1st
 strand cDNA was primed with oligo(dT)17 on 50 ng of
 DNase-treated, total cellular RNA obtained from 5,000-10
 ,000 microdissected preneoplastic cells
 histologically-determined to be prostatic intraepithelial
 neoplasia 2 (PIN2) cells. Double-stranded cDNA was
 ligated to EcoRI adaptors, 5 cycles of PCR applied to the
 cDNA with an adaptor-specific primer, and the resulting
 PCR product subcloned into pAMP10 by the UDG-cloning
 method (Life Technologies). Average insert size is 600
 bp. NOTE: Not directionally cloned. This library was
 constructed by David Krizman."
 54 a 20 c 30 g 46 t

BASE COUNT
 N
 Query Match 100.0%; Score 12; DB 10; Length 150;
 Best Local Similarity 100.0%; Pred. No. 7e+03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ggattttacagt 12
 |||
 Db 126 GGATTTTACAGT 115

RESULT 8
 AUI180504
 LOCUS AUI180504 Medaka eye cDNA library (SNK01) Oryzias latipes cDNA clone
 DEFINITION NGX10.02c, mRNA sequence.
 ACCESSION AUI180504
 VERSION AUI180504.1 GI:13429341
 KEYWORDS EST.
 SOURCE Japanese medaka.

ORGANISM

Oryzias latipes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
 Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.

REFERENCE 1 (bases 1 to 152)
 AUTHORS Sanaka,E., Hori,H., Naruse,K., Mitani,H. and Tanaka,M.
 TITLE Medaka EST analysis
 JOURNAL Unpublished (2001)
 COMMENT Contact: Emi Sanaka

Department of Biological Sciences
 Graduate School of Science, Nagoya University
 Furo-cho, Chikusa-ku, Nagoya 464-8602, Japan
 Tel: 81-52-789-2973
 Fax: 81-52-789-2974

Email: sanaka@bio.nagoya-u.ac.jp

This clone was isolated from Medaka eye cDNA library (SNK01) 5'end.

FEATURES

Location/Qualifiers
 1..152
 /organism="Oryzias latipes"
 /strain="wild type"
 /db_xref="taxon:8090"
 /clone="NGX10.02c"
 /clone_lib="Medaka eye cDNA library (SNK01)"
 /tissue_type="eye"
 /dev_stage="adult"
 /note="Wild samples from Okayama Pref. (Southern part of
 Japan)"
 56 a 23 c 30 g 43 t

BASE COUNT

ORIGIN
 Query Match 100.0%; Score 12; DB 10; Length 152;
 Best Local Similarity 100.0%; Pred. No. 7e+03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ggattttacagt 12
 |||
 Db 81 GGATTTTACAGT 92

RESULT

9
 AQQ09485 155 bp DNA GSS 27-JUN-1998
 LOCUS CIT-HSP-2283N16.TRB CIT-HSP Homo sapiens genomic clone 2283N16, DNA
 DEFINITION sequence.
 ACCESSION AQQ09485
 VERSION AQQ09485.1 GI:3128870
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 155)
 AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
 Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
 Simon,M. and Venter,J.C.
 TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map
 Building (1998)

Unpublished (1998)

Other_GSSs: CIT-HSP-2283N16.TFB

Contact: Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: mdadams@tigr.org

Clones are available from Research Genetics (info@resgen.com). BAC
 end search page:

http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html

Seq primer: M13 Reverse

Class: BAC ends.


```

FEATURES             Location/Qualifiers
  source              1..155
                     /organism="Homo sapiens"
                     /db_xref="GDB:7148131"
                     /db_xref="taxon:9606"
                     /clone="2283N16"
                     /clone_lib="CIT-HSP"
                     /sex="Male"
                     /cell_type="Sperm"
                     /note="Vector: pBelOBAcl1; Site_1: HindIII; Site_2: HindIII"
BASE COUNT          30 a 46 c 34 g 45 t
ORIGIN
Query Match          100.0%; Score 12; DB 13; Length 155;
Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggattttacagt 12
    |||||
Db 11 GGATTTTACAGT 22

RESULT 10
AV333933
LOCUS
DEFINITION
MUS musculus cDNA clone 6330552F17 3', similar to X04070 Rat liver
mRNA for gap junction protein, mRNA sequence.
AV333933
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 165)
Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai
,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata
,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H.,
Suzuki,H., Takahashi,F., Tateno,M., Tominaga,N., Tsunoda,Y.,
Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T.,
Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Konno,H., et al. 1999)
Unpublished (1999)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
Matsura,S., Carninci,P., Akizawa,J., Shibata,K., Izawa,M., Kawai,J.,
Y., and Hayashizaki,Y.
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
Location/Qualifiers
  source              1..165
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="6330552F17"
                     /clone_lib="RIKEN full-length enriched, adult male medulla
                     oblongata"
                     /sex="male"
                     /tissue_type="medulla oblongata"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGGATCCAGAGCTCTTTTCTTTTCTTTTCTTTTCTTTN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 100.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGATTCTCGAGTTAATTAATTAATCCCCCCCCCC
                     3']. cDNA was cloned into the XhoI and BamHI sites.
                     Vector: a modified phuescript KS(+) after bulk excision
                     from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
                     BamHI"
BASE COUNT          44 a 31 c 46 g 44 t
ORIGIN
Query Match          100.0%; Score 12; DB 10; Length 165;
Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggattttacagt 12
    |||||
Db 153 GGATTTTACAGT 164

RESULT 11
BF881282
LOCUS
DEFINITION
QVI-ET0183-021200-529-q05_1 ET0183 Homo sapiens cDNA, mRNA
sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 165)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Negai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922

```

Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?l1-QV1et2-QV1-ET0183-021200-529-d05-1st3-2000-12-02&t4=1)
 Seq primer: puc 18 forward
 High quality sequence stop: 165.

FEATURES

source

```

1. 165
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="ET0183"
/dev_stage="Adult"
/note="Organ: lung_tumor; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT      34 a 46 c 35 g 50 t
ORIGIN

```

```

Query Match      100.0%; Score 12; DB 11; Length 165;
-Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy: 1 ggattttacagt 12
      |||||
Db: 96 GGATTTTACAGT 107

```

```

RESULT 12
BF909645
LOCUS      BF909645      171 bp      mRNA      EST      18-JAN-2001
DEFINITION PM3-UT0058-181000-007-g01 UT0058 Homo sapiens cDNA, mRNA sequence.
ACCESSION  BF909645
VERSION     BF909645.1 GI:12301103
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens

```

```

REFERENCE
AUTHORS   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 171)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

```

```

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l1=PM3st2-PM3-UT0058-181000-007-g01&t3=2000-10-18&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 17
High quality sequence stop: 113.
Location/Qualifiers
1. .171

```

FEATURES
 source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="UT0058"
/dev_stage="Adult"
/note="Organ: uterus_tumor; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT      34 a 42 c 49 g 46 t
ORIGIN

```

```

Query Match      100.0%; Score 12; DB 11; Length 171;
-Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy: 1 ggattttacagt 12
      |||||
Db: 117 GGATTTTACAGT 128

```

```

RESULT 13
AI029683/c
LOCUS      AI029683      172 bp      mRNA      EST      04-JUL-1999
DEFINITION UI-R-CO-1y-b-01-0-UI.s1 UI-R-CO Rattus norvegicus cDNA clone
ACCESSION  AI029683
VERSION     AI029683.1 GI:4300245
KEYWORDS   EST.
SOURCE     Norway rat.
ORGANISM   Rattus norvegicus

```

```

REFERENCE
AUTHORS   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 172)
Bonaldi,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene discovery
Genome Res. 6 (9), 791-806 (1996)
9704477
On Jun 22, 1998 this sequence version replaced gi:3247509.
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu

```

```

The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized adult liver library. cDNA Library Preparation: M. Fatima Bonaldi, Ph.D. Clone distribution: clones will be available through Research Genetics. This clone is also available through the I.M.A.G.E. Consortium at LLNL (info@image.llnl.gov). IMAGE ID=1782730
Seq primer: M13 Forward
POLYA=No.
Location/Qualifiers
1. .172
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CO-1y-b-01-0-UI"
/clone_lib="UI-R-CO"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CO library is a subtracted library derived from the UI-R-A1 and UI-R-E1 libraries. The UI-R-A1 library consisted of a mixture of individually tagged normalized libraries

```

FEATURES
 source

constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, and muscle. The UI-R-EI library consisted of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dr track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-CO) was constructed as follows: PCR amplified cDNA inserts from a pool of UI-R-AI and UI-R-EI clones from which 3' ESTs had been derived was used as a driver in a hybridization with the pooled UI-R-AI and UI-R-EI library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-CO library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)"

BASE COUNT 32 a 47 c 32 g 61 t
ORIGIN

Query Match 100.0%; Score 12; DB 10; Length 172;
Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggattttacagt 12
|||||
Db 29 GGATTTTACAGT 18

RESULT 14
A2818519 173 bp DNA GSS 20-FEB-2001
LOCUS
DEFINITION 2M0088A10R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC2M0088A10 R, DNA sequence.

ACCESSION A2818519
VERSION A2818519.1 GI:12988427
KEYWORDS GSS.
SOURCE house mouse.

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 173)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0088 row: A column: 10
Seq primer: CACACGGAACACTATGACC
Class: plasmid ends
High quality sequence stop: 173.

FEATURES
source
1. .173
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0088A10"
/clone_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (gil4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 41 a 41 c 46 g 45 t
ORIGIN

Query Match 100.0%; Score 12; DB 13; Length 173;
Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggattttacagt 12
|||||
Db 134 GGATTTTACAGT 145

RESULT 15
A0985182/c 177 bp DNA GSS 30-JAN-2000
LOCUS
DEFINITION RPCI-23-307M5.TJ RPCI-23 Mus musculus genomic clone RPCI-23-307M5, DNA sequence.

ACCESSION A0985182
VERSION A0985182.1 GI:6818387
KEYWORDS GSS.
SOURCE house mouse.

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 177)
Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P. and Fraser,C.M.

TITLE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: RPCI-23-307M5.TV

Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@jeong.med.bufo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.bufo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac-ends/mouse/bac_end_intro.html
Plate: 307 row: M column: 5
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1. .177
/organism="Mus musculus"
/strain="C57BL/6J"

```

/db_xref="taxon:10090"
/clone="RPCI-23-307M5"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACe3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies). "
```

```

BASE COUNT      56 a      48 c      20 g      52 t      1 others
ORIGIN

Query Match      100.0%; Score 12; DB 13; Length 177;
Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ggattttacagt 12
        |||
        139 GGATTTTACAGT 128
```

Search completed: April 3, 2002, 04:39:12
Job time: 2832 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2002, 03:52:00 ; Search time 2120.21 Seconds
(without alignments)
60.819 Million cell updates/sec

Title: us-09-700-187-1

Perfect score: 12

Sequence: 1 ggattttacagt 12

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

- 1: em_estfun:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estom:*
- 5: em_estpl:*
- 6: em_estba:*
- 7: em_estov:*
- 8: em_hic:*
- 9: em_hic:*
- 10: gb_estl:*
- 11: gb_est2:*
- 12: gb_hic:*
- 13: gb_gss:*
- 14: em_gss_fun:*
- 15: em_gss_hum:*
- 16: em_gss_inv:*
- 17: em_gss_pln:*
- 18: em_gss_pro:*
- 19: em_gss_rod:*
- 20: em_gss_vrt:*
- 21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	12	100.0	88	13	AZ925237 4910.ez32
C 2	12	100.0	118	10	AW217731 EST296445
C 3	12	100.0	126	13	AZ289196 RPI-23-5
C 4	12	100.0	130	10	A1610647 tp20402.x
C 5	12	100.0	147	10	BE145737 IL5-HT020
C 6	12	100.0	147	11	D25785 HUMGS04153
C 7	12	100.0	150	10	AA579315 n136806.S
C 8	12	100.0	152	10	AU180504 AU180504
C 9	12	100.0	153	13	AQ009485 CIT-HSP-2
C 10	12	100.0	165	10	AV333933 AV333933
C 11	12	100.0	165	11	BF881282 QV1-ET018
C 12	12	100.0	171	11	BF909645 PM3-UT005

C 13	12	100.0	172	10	AI029683
C 14	12	100.0	173	13	AZ818519
C 15	12	100.0	177	13	AQ985182
C 16	12	100.0	180	11	BF235518
C 17	12	100.0	184	10	AV054690
C 18	12	100.0	185	10	AV071084
C 19	12	100.0	185	10	BB357868
C 20	12	100.0	185	11	BG942699
C 21	12	100.0	186	10	AI345298
C 22	12	100.0	189	11	F29452
C 23	12	100.0	190	10	AV288772
C 24	12	100.0	193	10	AV285178
C 25	12	100.0	194	10	AI136924
C 26	12	100.0	202	11	D60959
C 27	12	100.0	207	10	AI572468
C 28	12	100.0	210	10	AV336005
C 29	12	100.0	216	13	B25087
C 30	12	100.0	217	10	AI563911
C 31	12	100.0	220	13	AQ278868
C 32	12	100.0	223	10	BE236288
C 33	12	100.0	223	10	BE236293
C 34	12	100.0	224	10	AV252038
C 35	12	100.0	224	10	BB512144
C 36	12	100.0	225	10	BB256670
C 37	12	100.0	226	11	BI220740
C 38	12	100.0	227	10	AA129692
C 39	12	100.0	227	11	BI293223
C 40	12	100.0	228	10	BB171532
C 41	12	100.0	228	10	BB310621
C 42	12	100.0	230	10	BB152264
C 43	12	100.0	231	10	AV368688
C 44	12	100.0	231	10	BB503511
C 45	12	100.0	231	10	BB510596

ALIGNMENTS

RESULT 1
AZ925237/c
LOCUS 88 bp DNA
DEFINITION 4910.ez32kl8.sl Saccharomyces paradoxus N17 Saccharomyces paradoxus genomic clone 4910.ez32kl8.sl, DNA sequence.
ACCESSION AZ925237
VERSION AZ925237.1 GI:13496136
KEYWORDS GSS:
SOURCE Saccharomyces paradoxus.
ORGANISM Saccharomyces paradoxus
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

REFERENCE 1 (bases 1 to 88)
AUTHORS Clifton, P.F., Hillier, L.W., Fulton, L., Graves, T., Miner, T., Gish, W.R., Waterston, R.H. and Johnston, M.
TITLE Surveying Saccharomyces genomes to identify functional elements by comparative DNA sequence analysis
JOURNAL Unpublished (2001)
COMMENT Contact: Johnston M
Department of Genetics
Washington University Medical School
Box 8232, 4566 Scott Ave., St. Louis, MO 63110, USA
Tel: 314 362 2735
Fax: 314 362 7855
Email: mj@genetics.wustl.edu
Class: random plasmid subclone.
Location/Qualifiers
1. .88
/organism="Saccharomyces paradoxus"
/strain="N17"
/db_xref="taxon:27291"
/clone="4910.ez32kl8.sl"
/clone_lib="Saccharomyces paradoxus N17"
/note="Random genomic sequence"

BASE COUNT 26 a 12 c 8 g 42 t

ORIGIN

Query Match 100.0%; Score 12; DB 13; Length 88;
 Best Local Similarity 100.0%; Pred. No. 7.3e+03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
 |||||

Db 40 GGATTTTACAGT 29

RESULT 2

AW217731/c
 LOCUS AW217731 118 bp mRNA EST 18-MAY-2001
 DEFINITION EST296445 tomato flower buds 8 mm to pre-anthesis, Cornell University Lycopersicon esculentum cDNA clone cT0C6E5, mRNA sequence.

ACCESSION AW217731.1 GI:6528605

VERSION AW217731

KEYWORDS EST.

SOURCE tomato.

ORGANISM

Lycopersicon esculentum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;

Lycopersicon.

1 (bases 1 to 118)

REFERENCE van der Hoeven, R.S., Bezzerides, J.L., Matern, A.L., Holt, I.E., Liang

AUTHORS F., Upton, J., Hansen, T., Craven, M.B., Bowman, C.L., Ahn, S., Ronning

C.M., Fraser, C.M., Martin, G.B., Giovannoni, J.J., and Tanksley, S.D.

Generation of ESTs from tomato flower tissue

Unpublished (1999)

Contact: CUGI

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>

5 prime sequence.

Location/Qualifiers

1..118

/organism="Lycopersicon esculentum"

/cultivar="TA496"

/db_xref="taxon:4081"

/clone="cT0C6E5"

/clone_lib="tomato flower buds 8 mm to pre-anthesis,

Cornell University"

/tissue_type="flower"

/dev_stage="buds 8mm-to-preanthesis"

/note="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:

XhoI; supplier: Tanksley; Flower buds and flowers were

taken from greenhouse plants (4-8 wks old, TA496). They

were immediately frozen in liquid nitrogen and then

size-separated while remaining frozen."

BASE COUNT 49 a 21 c 16 g 32 t

ORIGIN

Query Match 100.0%; Score 12; DB 10; Length 118;

Best Local Similarity 100.0%; Pred. No. 7.2e+03;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12

|||||

Db 47 GGATTTTACAGT 36

RESULT 3

AZ289196

LOCUS AZ289196 126 bp DNA GSS 27-JUL-2000

DEFINITION RPCI-23-59B23.TJB RPCI-23 Mus musculus genomic clone RPCI-23-59B23,

DNA sequence.

ACCESSION AZ289196

VERSION AZ289196.1 GI:9530982

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 126)

REFERENCE Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akimret

AUTHORS B., Levin, M., McGann, S., Tseydaye, G., Geer, K., Kroi, M., de Jong, P.

and Fraser, C.M.

Mouse BAC End Sequences from Library RPCI-23

Unpublished (1999)

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)

or from Resea ch Genetics (info@resgen.com). BAC end page:

http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html

Plate: 59 row: B column: 23

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

1..126

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-23-59B23"

/clone_lib="RPCI-23"

/sex="Female"

/lab_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:

EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or

brain genomic DNA was isolated and partially digested

with a combination of EcoRI and EcoRI Methylase. Size

selected DNA was cloned into the pBACe3.6 vector at the

EcoRI sites. The ligation products were transformed into

DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 38 a 20 c 30 g 38 t

ORIGIN

Query Match 100.0%; Score 12; DB 13; Length 126;

Best Local Similarity 100.0%; Pred. No. 7.1e+03;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12

|||||

Db 93 GGATTTTACAGT 104

RESULT 4

AI610647/c

LOCUS AI610647 130 bp mRNA EST 21-APR-1999

DEFINITION tp20402.x1 NCI-CGAP_Gas4 Homo sapiens cDNA clone IMAGE:2188323 3'

similar to SW:UIC6_HCMVA P16836 HYPOTHETICAL PROTEIN UL126.

:contains element LTR5 repetitive element ; , mRNA sequence.

ACCESSION AI610647

VERSION AI610647.1 GI:4619814

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 130)

REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

TITLE


```

RESULT 7
AA579315/c 150 bp mRNA EST 12-SEP-1997
LOCUS nf36e06.sl NCI_CGAP_Pr2 Homo sapiens cDNA clone IMAGE:915874, mRNA
DEFINITION
ACCESSION AA579315
VERSION AA579315.1 GI:2357499
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 150)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuqui, M.D.,
Michael Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Insert Length: 578 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 126.
FEATURES
source
1..150
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:915874"
/clone_lib="NCI_CGAP_Pr2"
/sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"
/note="vector: PAM10; Site_1: NotI; Site_2: EcoRI; 1st
strand cDNA was primed with oligo(dT)17 on 50 ng of
DNase-treated, total cellular RNA obtained from 5,000-10
,000 microdissected preneoplastic cells
histologically-determined to be prostatic intraepithelial
neoplasia 2 (PIN2) cells. Double-stranded cDNA was
ligated to EcoRI adaptors, 5 cycles of PCR applied to the
cDNA with an adaptor-specific primer, and the resulting
PCR product subcloned into pAMP10 by the UDG-cloning
method (Life Technologies). Average insert size is 600
bp. NOTE: Not directionally cloned. This library was
constructed by David Krizman."
BASE COUNT 54 a 20 c 30 g 46 t
ORIGIN
1 ggaatttcacgt 12
|||||
Db 126 GGATTTTACAGT 115

Query Match 100.0%; Score 12; DB 10; Length 150;
Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ggaatttcacgt 12
|||||
Db 81 GGATTTTACAGT 92

RESULT 8
AA180504 152 bp mRNA EST 21-MAR-2001
LOCUS AA180504 Medaka eye cDNA library (SNK01) Oryzias latipes cDNA clone
DEFINITION NCY10.02c, mRNA sequence.
ACCESSION AA180504
VERSION AA180504.1 GI:13429341
KEYWORDS EST.
SOURCE Japanese medaka.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 152)
AUTHORS Sanaka,E., Hori,H., Naruse,K., Mitani,H. and Tanaka,M.
TITLE Medaka EST analysis
JOURNAL Unpublished (2001)
COMMENT Contact: Emi Sanaka
Department of Biological Sciences
Graduate School of Science, Nagoya University
Furo-cho, Chikusa-ku, Nagoya 464-8602, Japan
Tel: 81-52-789-2973
Fax: 81-52-789-2974
Email: sanaka@bio.nagoya-u.ac.jp
This clone was isolated from Medaka eye cDNA library (SNK01) 5'end.
FEATURES
source
1..152
/organism="Oryzias latipes"
/strain="wild type"
/db_xref="taxon:8090"
/clone="NGY10.02c"
/clone_lib="Medaka eye cDNA library (SNK01)"
/tissue_type="eye"
/dev_stage="adult"
/note="Wild samples from Okayama Pref.(Southern part of
Japan)"
BASE COUNT 56 a 23 c 30 g 43 t
ORIGIN
1 ggaatttcacgt 12
|||||
Db 81 GGATTTTACAGT 92

Query Match 100.0%; Score 12; DB 10; Length 152;
Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ggaatttcacgt 12
|||||
Db 81 GGATTTTACAGT 92

RESULT 9
AA009485 155 bp DNA GSS 27-JUN-1998
LOCUS CIT-HSP-2283N16.TRB CIT-HSP Homo sapiens genomic clone 2283N16, DNA
DEFINITION
ACCESSION AA009485
VERSION AA009485.1 GI:3128870
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 155)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
Simon,M. and Venter,J.C.
TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
JOURNAL Unpublished (1998)
COMMENT Other_GSSs: CIT-HSP-2283N16.TFB
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: M13 Reverse
Class: BAC ends.

```

```

ORGANISM Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE 1 (bases 1 to 152)
AUTHORS Sanaka,E., Hori,H., Naruse,K., Mitani,H. and Tanaka,M.
TITLE Medaka EST analysis
JOURNAL Unpublished (2001)
COMMENT Contact: Emi Sanaka
Department of Biological Sciences
Graduate School of Science, Nagoya University
Furo-cho, Chikusa-ku, Nagoya 464-8602, Japan
Tel: 81-52-789-2973
Fax: 81-52-789-2974
Email: sanaka@bio.nagoya-u.ac.jp
This clone was isolated from Medaka eye cDNA library (SNK01) 5'end.
FEATURES
source
1..152
/organism="Oryzias latipes"
/strain="wild type"
/db_xref="taxon:8090"
/clone="NGY10.02c"
/clone_lib="Medaka eye cDNA library (SNK01)"
/tissue_type="eye"
/dev_stage="adult"
/note="Wild samples from Okayama Pref.(Southern part of
Japan)"
BASE COUNT 56 a 23 c 30 g 43 t
ORIGIN
1 ggaatttcacgt 12
|||||
Db 81 GGATTTTACAGT 92

Query Match 100.0%; Score 12; DB 10; Length 152;
Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ggaatttcacgt 12
|||||
Db 81 GGATTTTACAGT 92

RESULT 9
AA009485 155 bp DNA GSS 27-JUN-1998
LOCUS CIT-HSP-2283N16.TRB CIT-HSP Homo sapiens genomic clone 2283N16, DNA
DEFINITION
ACCESSION AA009485
VERSION AA009485.1 GI:3128870
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 155)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
Simon,M. and Venter,J.C.
TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
JOURNAL Unpublished (1998)
COMMENT Other_GSSs: CIT-HSP-2283N16.TFB
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: M13 Reverse
Class: BAC ends.

```



```

FEATURES             Location/Qualifiers
  source              1..155
                    /organism="Homo sapiens"
                    /db_xref="GDB:7148131"
                    /db_xref="taxon:9606"
                    /clone="2283N16"
                    /clone_lib="CIT-HSP"
                    /sex="Male"
                    /cell_type="Sperm"
                    /note="Vector: pBelOBAC11; Site_1: HindIII; Site_2: HindIII"
BASE COUNT          30 a 46 c 34 g 45 t
ORIGIN
1 ggattttacagt 12
|||||
11 GGATTTTACAGT 22

Query Match          100.0%; Score 12; DB 13; Length 155;
Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
|||||
11 GGATTTTACAGT 22

RESULT 10
AV333933
LOCUS               165 bp      mRNA               11-NOV-1999
DEFINITION          AV333933 RIKEN full-length enriched, adult male medulla oblongata
                    Mus musculus cDNA clone 6330552F17 3' similar to X04070 Rat liver
                    mRNA for gap junction protein, mRNA sequence.
ACCESSION            AV333933
VERSION              AV333933.1 GI:6373985
KEYWORDS             house mouse.
SOURCE               house mouse.
ORGANISM             Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 165)
Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai
,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata
,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H.,
Suzuki,H., Takahashi,F., Taleno,M., Tomimaga,N., Tsunoda,Y.,
Watabiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T.,
Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Konno,H., et al. 1999)
Unpublished (1999)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Sasaki,N., Izawa,M., Watabiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
Matsura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki
,Y.
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
Itoh,M., Kitzunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for

FEATURES             Location/Qualifiers
  source              1..165
                    /organism="Mus musculus"
                    /strain="C57BL/6J"
                    /db_xref="taxon:10090"
                    /clone="6330552F17"
                    /clone_lib="RIKEN full-length enriched, adult male medulla
                    oblongata"
                    /sex="male"
                    /tissue_type="medulla oblongata"
                    /dev_stage="adult"
                    /lab_host="DH10B"
                    /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                    prepared and sequenced in Mouse Genome Encyclopedia
                    Project of Genome Exploration Research Group in Riken
                    Genomic Sciences Center and Genome Science Laboratory in
                    RIKEN. Division of Experimental Animal Research in Riken
                    contributed to prepare mouse tissues. 1st strand cDNA was
                    primed with a primer 15',
                    GAGAGAGAGAGATCCACAGCTCTTTTCTTTTCTTNN 3', cDNA was
                    prepared by using trehalose thermo-activated reverse
                    transcriptase and subsequently enriched for full-length by
                    cap-trapper. cDNA went through one round of normalization
                    to Rot = 10.0 and subtraction to Rot = 100.0. Second
                    strand cDNA was prepared with the primer adapter of
                    sequence 15' GAGAGAGAGATTCGAGTTAAATTAATATCCCTCCCTCCCTCC
                    3'. cDNA was cloned into the XhoI and BamHI sites.
                    Vector: a modified pBluescript KS(+) after bulk excision
                    from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
                    BamHI"
BASE COUNT          44 a 31 c 46 g 44 t
ORIGIN
1 ggattttacagt 12
|||||
11 GGATTTTACAGT 164

Query Match          100.0%; Score 12; DB 10; Length 165;
Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
|||||
11 GGATTTTACAGT 164

RESULT 11
BF881282
LOCUS               165 bp      mRNA               17-JAN-2001
DEFINITION          QV1-ET0183-021200-529-d05_1 ET0183 Homo sapiens cDNA, mRNA
                    sequence.
ACCESSION            BF881282
VERSION              BF881282.1 GI:12271408
KEYWORDS             EST.
SOURCE               human.
ORGANISM             Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Calarrhini; Hominidae; Homo.
1 (bases 1 to 165)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.P.,
Goldman,G.H., Matsumoto,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.K., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922

```

Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV1&t2=QV1-ET0183-021200-529-005-1&t3=2000-12-02&t4=1>)
 Seq primer: puc 18 forward
 High quality sequence stop: 165.

FEATURES

Location/Qualifiers
 1..165
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="ET0183"
 /dev_stage="Adult"
 /note="Organ: lung_tumor; Vector: puc18; Site_1: Smal;
 Site_2: Smal; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the puc 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."

BASE COUNT 34 a 46 c 35 g 50 t
 ORIGIN
 1 ggaattaccagt 12
 |||||
 Db 96 GGATTTTACAGT 107

Query Match 100.0%; Score 12; DB 11; Length 165;
 -Best Local Similarity 100.0%; Pred. No. 7e+03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ggaattaccagt 12
 |||||
 Db 96 GGATTTTACAGT 107

RESULT 12
 BF909645
 LOCUS PM3-UT0058-181000-007-g01 UT0058 Homo sapiens cDNA, mRNA sequence.
 DEFINITION BF909645
 ACCESSION BF909645
 VERSION BF909645.1 GI:12301103
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 171)
 AUTHORS Dias Neto E., Garcia Correa R., Verjovski-Almeida S., Briones M.R.,
 Nagai M.A., da Silva W. Jr., Zago M.A., Bordin S., Costa F.F.,
 Goldman G.H., Carvalho A.F., Matsukuma A., Baia G.S., Simpson D.H.,
 Brumstein A., de Oliveira P.S., Bucher P., Jongeneel C.V., O'Hare
 M.J., Soares F., Brentani R.R., Reis L.F., de Souza S.J. and
 Simpson A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663

COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001

Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM3&t2=PM3-UT0058-181000-007-g01&t3=2000-10-18&t4=1>)
 Seq primer: puc 18 forward
 High quality sequence start: 17
 High quality sequence stop: 113.

FEATURES

Location/Qualifiers
 1..171

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="UT0058"
 /dev_stage="Adult"
 /note="Organ: uterus_tumor; Vector: puc18; Site_1: Smal;
 Site_2: Smal; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the puc 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."

BASE COUNT 34 a 42 c 49 g 46 t
 ORIGIN
 1 ggaattaccagt 12
 |||||
 Db 117 GGATTTTACAGT 128

Query Match 100.0%; Score 12; DB 11; Length 171;
 -Best Local Similarity 100.0%; Pred. No. 7e+03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ggaattaccagt 12
 |||||
 Db 117 GGATTTTACAGT 128

RESULT 13
 AI029683/c
 LOCUS UI-R-CO-iy-b-01-0-UI-s1 UI-R-CO Rattus norvegicus cDNA clone
 DEFINITION UI-R-CO-iy-b-01-0-UI 3', mRNA sequence.
 ACCESSION AI029683
 VERSION AI029683.1 GI:4300245
 KEYWORDS EST.
 SOURCE Norway rat.

ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 REFERENCE 1 (bases 1 to 172)
 AUTHORS Bonaldo M.F., Lennon G. and Soares M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477

COMMENT On Jun 22, 1998 this sequence version replaced gi:3247509.

Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu

The sequence tag present in the cDNA between the NotI site and the
 oligo-dT track served to identify it as a clone from the normalized
 adult Liver library. cDNA Library Preparation: M. Fatima Bonaldo,
 Ph.D. Clone distribution: clones will be available through Research
 Genetics This clone is also available through the I.M.A.G.E.
 Consortium at LNL (info@image.llnl.gov). IMAGE ID=1782730
 Seq primer: M13 Forward
 POLYA=No.

FEATURES

Location/Qualifiers
 1..172
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-CO-iy-b-01-0-UI"
 /clone_lib="UI-R-CO"
 /dev_stage="adult"
 /lab_host="Dhl08 (Life Technologies)"

/note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CO
 library is a subtracted library derived from the UI-R-A1
 and UI-R-E1 libraries. The UI-R-A1 library consisted of a
 mixture of individually tagged normalized libraries

constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, and muscle. The UI-R-E1 library consisted of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryos. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-C0) was constructed as follows: PCR amplified cDNA inserts from a pool of UI-R-A1 and UI-R-E1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the pooled UI-R-A1 and UI-R-E1 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-C0 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996).

BASE COUNT 32 a 47 c 32 g 61 t
ORIGIN

Query Match 100.0%; Score 12; DB 10; Length 172;
Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gattttacagt 12
|||||

DB 29 GGATTTTACAGT 18

RESULT 14

AZ818519 173 bp DNA GSS 20-FEB-2001
LOCUS 2M0088A10R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION clone UUGC2M0088A10 R, DNA sequence.
ACCESSION AZ818519
VERSION A2818519.1 GI:12988427
KEYWORDS GSS
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 173)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A., and Wright, D., Weiss, R.
REFERENCE AUTHORS

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0088 row: A column: 10
Seq primer: CACACAGGAGACGTATGACC
Class: plasmid ends
High quality sequence stop: 173.
Location/Qualifiers
1. 173
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0088A10"
/clone_lib="Mouse 10kb plasmid UUGC1M library"

FEATURES

SOURCE

1. 173
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0088A10"
/clone_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 41 a 41 c 46 g 45 t
ORIGIN

Query Match 100.0%; Score 12; DB 13; Length 173;
Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gattttacagt 12
|||||

DB 134 GGATTTTACAGT 145

RESULT 15

AQ985182/c 177 bp DNA GSS 30-JAN-2000
LOCUS AQ985182 TJ RPCI-23 Mus musculus genomic clone RPCI-23-307M5, DNA sequence.
ACCESSION AQ985182
VERSION AQ985182.1 GI:6818387
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 177)
Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Kroll, M., de Jong, P., and Fraser, C. M.
REFERENCE AUTHORS

TITLE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: RPCI-23-307M5.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 307 row: M column: 5
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1. 177
/organism="Mus musculus"
/strain="C57BL/6J"

FEATURES

SOURCE

1. 177
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"

/db_xref="taxon:10090"
/clone="RPCI-23-307M5"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACe3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT 56 a 48 C 20 g 52 t 1 others
-ORIGIN

Query Match 100.0%; Score 12; DB 13; Length 177;
Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
UY 1 ggattttacagt 12
| | | | | | | | | |
Db 139 GGATTTTACAGT 128

Search completed: April 3, 2002, 04:39:12
Job time: 2832 sec

US-08-998-416-1071

Query Match 100.0%; Score 12; DB 4; Length 679;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
|||||
Db 236 GGATTTTACAGT 225

RESULT 2

US-08-993-228-5/c
; Sequence 5, Application US/08993228
; Patent No. 5976838

; GENERAL INFORMATION:

; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: LaVallie, Edward R.
; APPLICANT: Racie, Lisa A.
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; APPLICANT: Agostino, Michael J.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02140

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/993,228
; FILING DATE:
; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Sprunger, Suzanne A.
; REGISTRATION NUMBER: 41,323

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617) 498-8284
; TELEFAX: (617) 876-5851

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1138 base pairs
; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

US-08-993-228-5

Query Match 100.0%; Score 12; DB 2; Length 1138;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
|||||
Db 985 GGATTTTACAGT 974

RESULT 3

US-08-903-801-2/c
; Sequence 2, Application US/08903801
; Patent No. 5932712

; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: NEW ANEXIN BINDING PROTEIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/903,801
; FILING DATE: Herewith
; CLASSIFICATION: 536

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0354 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1434 base pairs
; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: PROSNON01

; CLONE: 2272281

US-08-903-801-2

Query Match 100.0%; Score 12; DB 2; Length 1434;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
|||||
Db 1002 GGATTTTACAGT 991

RESULT 4

US-09-295-055-2/c
; Sequence 2, Application US/09295055
; Patent No. 6232440

; GENERAL INFORMATION:

; APPLICANT: Hillman, Jennifer
; APPLICANT: Corley, Neil C.

; APPLICANT: Shah, Purvi

; TITLE OF INVENTION: NEW ANEXIN BINDING PROTEIN

; NUMBER OF SEQUENCES: 3

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/295,055
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/903,801
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PP-0354 US
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1434 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PROSNON01
; CLONE: 2272281
; US-09-295-055-2

Query Match 100.0%; Score 12; DB 4; Length 1434;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ggattttacagt 12
|||||

Db 1002 GGATTTTACAGT 991

RESULT 5
US-07-828-700-7
Sequence 7, Application US/07828700
Patent No. 5308760

; GENERAL INFORMATION:
; APPLICANT: Brown, Kit L.
; TITLE OF INVENTION: CRYSTAL PROTEINS OF BACILLUS
; TITLE OF INVENTION: THURINGIENSIS, GENES ENCODING THEM, AND HOSTS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wegner, Cantor, Mueller & Player
; STREET: 1233 20th Street, N.W.
; CITY: Washington
; STATE: D.C.
; ZIP: 20036-8218

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/828,700
; FILING DATE: 19920203
; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/817,915
; FILING DATE: 10 JAN 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Cantor, Herbert I.
; REGISTRATION NUMBER: 24,392
; REFERENCE/DOCKET NUMBER: P-9575-23207
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-887-0400
; TELEFAX: 202-835-0605
; TELEX: 440706
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2259 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Bacillus thuringiensis
; STRAIN: thompson1
; US-07-828-700-7

Query Match 100.0%; Score 12; DB 1; Length 2259;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ggattttacagt 12
|||||

Db 1536 GGATTTTACAGT 1547

RESULT 6
US-08-343-760A-1
Sequence 1, Application US/08343760A
Patent No. 5679783

; GENERAL INFORMATION:
; APPLICANT: De Robertis, Edward M
; APPLICANT: Sasal, Yoshiaki
; TITLE OF INVENTION: Tissue Differentiation Affecting
; TITLE OF INVENTION: Factor and Composition
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Majestic, Parsons, Siebert & Hsue
; STREET: Four Embarcadero Center, Suite 1450
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/343,760A
; FILING DATE: 22-NOV-1994
; CLASSIFICATION: 536

; ATTORNEY/AGENT INFORMATION:
; NAME: Siebert, J. Suzanne
; REGISTRATION NUMBER: 28,758
; REFERENCE/DOCKET NUMBER: 3100.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 363-5556
; TELEFAX: (415) 362-5418
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3796 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-08-343-760A-1

Query Match 100.0%; Score 12; DB 1; Length 3796;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ggattttacagt 12

```
Db 3249 GGATTTACAGT 3260
|||||
RESULT 7
US-08-061-697-33/c
; Sequence 33, Application US/08061697
; Patent No. 5498696
; GENERAL INFORMATION:
; APPLICANT: Brown, Michael S.; Briggs, Michael R.; Wang,
; APPLICANT: Xiaodong; Goldstein, Joseph L.
; TITLE OF INVENTION: Sterol Regulatory Element Binding Proteins
; TITLE OF INVENTION: and Their Use in Screening Assays
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P. O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/061,697
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: UTSD:347/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 320-7200
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 38 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-061-697-34

Query Match 91.7%; Score 11; DB 1; Length 38;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gattttacagt 12
|||||
Db 16 GATTTTACAGT 26

RESULT 9
US-08-131-365B-33/c
; Sequence 33, Application US/08131365B
; Patent No. 5527690
; GENERAL INFORMATION:
; APPLICANT: Brown, Michael S.
; APPLICANT: Briggs, Michael R.
; APPLICANT: Wang, Xiaodong
; APPLICANT: Goldstein, Joseph L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING
; TITLE OF INVENTION: TO STEROL REGULATORY ELEMENT BINDING
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P. O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/131,365B
; FILING DATE: 01-OCT-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: UTSD:372/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:

Db 3249 GGATTTACAGT 3260
|||||
RESULT 8
US-08-061-697-34
; Sequence 34, Application US/08061697
; Patent No. 5498696
; GENERAL INFORMATION:
; APPLICANT: Brown, Michael S.; Briggs, Michael R.; Wang,
; APPLICANT: Xiaodong; Goldstein, Joseph L.
; TITLE OF INVENTION: Sterol Regulatory Element Binding Proteins
; TITLE OF INVENTION: and Their Use in Screening Assays
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P. O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210

Query Match 91.7%; Score 11; DB 1; Length 38;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gattttacagt 12
|||||
Db 27 GATTTTACAGT 17
```


;
; LENGTH: 38 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "DNA"
; US-08-131-365B-33

Query Match 91.7%; Score 11; DB 1; Length 38;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 gattttacagt 12
| | | | | | | | | |
Db 27 GATTTTACAGT 17

RESULT 10
US-08-131-365B-34
; Sequence 34, Application US/08131365B
; Patent No. 5527690
; GENERAL INFORMATION:
; APPLICANT: Brown, Michael S.
; APPLICANT: Briggs, Michael R.
; APPLICANT: Wang, Xiaodong
; APPLICANT: Goldstein, Joseph L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING
; TO STEROL REGULATORY ELEMENT BINDING
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/131,365B
; FILING DATE: 01-OCT-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: UTSD:372/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 474-7577
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 38 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "DNA"
; US-08-131-365B-34

Query Match 91.7%; Score 11; DB 1; Length 38;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 gattttacagt 12
| | | | | | | | | |
Db 16 GATTTTACAGT 26

RESULT 11
US-08-668-123-33/C
; Sequence 33, Application US/08668123
; Patent No. 5891631
; GENERAL INFORMATION:
; APPLICANT: Brown, Michael S.
; APPLICANT: Briggs, Michael R.
; APPLICANT: Wang, Xiaodong
; APPLICANT: Goldstein, Joseph L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING
; TO STEROL REGULATORY ELEMENT BINDING
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/668,123
; FILING DATE: 14-JUN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/131,365
; FILING DATE: 01-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: UTSD:372/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 38 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "DNA"
; US-08-668-123-33

Query Match 91.7%; Score 11; DB 2; Length 38;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 gattttacagt 12
| | | | | | | | | |
Db 27 GATTTTACAGT 17

RESULT 12
US-08-668-123-34
; Sequence 34, Application US/08668123
; Patent No. 5891631
; GENERAL INFORMATION:
; APPLICANT: Brown, Michael S.
; APPLICANT: Briggs, Michael R.
; APPLICANT: Wang, Xiaodong
; APPLICANT: Goldstein, Joseph L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING
; TO STEROL REGULATORY ELEMENT BINDING
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 64

```
;
; CORRESPONDENCE ADDRESS:
; ADDRESS: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/668,123
; FILING DATE: 14-JUN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/131,365
; FILING DATE: 01-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: UTSD:372/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 38 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "DNA"
;
US-08-668-123-34

Query Match          91.7%; Score 11; DB 2; Length 38;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 gattttacagt 12
Db 16 GATTTTACAGT 26

RESULT 13
US-07-748-761-1/C
; Sequence 1, Application US/07748761
; Patent No. 5304490
; GENERAL INFORMATION:
; APPLICANT: Grierson, Donald
; APPLICANT: Pray, Rupert G
; APPLICANT: Lycett, Grantley W
; APPLICANT: Bird, Colin R
; APPLICANT: Ray, John A
; APPLICANT: Schuch, Wolfgang W
; TITLE OF INVENTION: DNA, DNA constructs, cells and plants
; TITLE OF INVENTION: derived therefrom
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cushman, Darby and Cushman
; STREET: Eleventh floor, 1615 L Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036-5601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
```

```
;
; APPLICATION NUMBER: US/07748,761
; FILING DATE: 19910823
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9018612.3
; FILING DATE: 24-AUG-1990
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 332 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
;
US-07-748-761-1

Query Match          91.7%; Score 11; DB 1; Length 332;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggattttacag 11
Db 122 GGATTTTACAG 112

RESULT 14
US-08-470-720-3/C
; Sequence 3, Application US/08470720
; Patent No. 5824798
; GENERAL INFORMATION:
; APPLICANT: HOFVANDER, Per
; APPLICANT: PERSSON, Per T
; APPLICANT: WIKSTROM, Olle
; APPLICANT: TALLBERG, Anneli
; TITLE OF INVENTION: GENETICALLY ENGINEERED MODIFICATION OF
; TITLE OF INVENTION: POTATO TO FORM AMYLOPECTIN-TYPE STARCH
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,720
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/070,455
; FILING DATE: 09-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Crane-Feury, Sharon E
; REGISTRATION NUMBER: 36,113
; REFERENCE/DOCKET NUMBER: 003300-293
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 492 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
```

; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..15
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 101..218
US-08-470-720-3

Query Match 91.7%; Score 11; DB 1; Length 492;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggattttacag 11
|||||
Db 367 GGATTTTACAG 357

RESULT 15
US-09-318-661-3/c
; Sequence 3, Application US/09318661
; Patent No. 6268488
; GENERAL INFORMATION:
; APPLICANT: Barbas III, Carlos F.
; APPLICANT: Shabat, Doron
; APPLICANT: Rader, Christoph
; APPLICANT: List, Benjamin
; APPLICANT: Lerner, Richard A.
; TITLE OF INVENTION: PRODRUG ACTIVATION USING CATALYTIC ANTIBODIES
; FILE REFERENCE: PLF00115
; CURRENT APPLICATION NUMBER: US/09/318,661
; CURRENT FILING DATE: 1999-05-25
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 856
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Nucleotide
; OTHER INFORMATION: residue sequence of catalytic fragment.
; NAME/KEY: CDS
; LOCATION: (1)..(855)
US-09-318-661-3

Query Match 91.7%; Score 11; DB 4; Length 856;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 gattttacagt 12
|||||
Db 759 GATTTTACAGT 749

Search completed: April 3, 2002, 05:14:51
Job time: 3901 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2002, 05:14:06 ; Search time 1999.71 Seconds
(without alignments)
19180.752 Million cell updates/sec

Title: US-09-700-187-3

Perfect score: 2325

Sequence: 1 agccttaagcgcaaggaa.....ttttgatcttgacaagaa 2325

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

1: gb_ba:**

2: gb_htg:**

3: gb_in:**

4: gb_om:**

5: gb_ov:**

6: gb_pat:**

7: gb_ph:**

8: gb_pl:**

9: gb_pr:**

10: gb_ro:**

11: gb_sts:**

12: gb_sy:**

13: gb_un:**

14: gb_vi:**

15: em_ba:**

16: em_fun:**

17: em_hum:**

18: em_in:**

19: em_om:**

20: em_or:**

21: em_ov:**

22: em_pat:**

23: em_ph:**

24: em_pl:**

25: em_ro:**

26: em_sts:**

27: em_sy:**

28: em_un:**

29: em_vi:**

30: em_htgo_hum:**

31: em_htgo_inv:**

32: em_htgo_rod:**

33: em_htg_hum:**

34: em_htg_inv:**

35: em_htg_rod:**

36: em_htg_other:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2325	100.0	3441	8	AB007911	AB007911 Pism sat
2	100	4.3	20785	8	AF138704	AF138704 Pism sat
3	89	3.8	5785	8	PSBLOX12	X78580 P.sativum (
4	82.4	3.5	326924	2	AC093082	Y17468 Pism sativ
5	82	3.5	1350	8	PSA17468	AC005083 Homo sapi
6	72.4	3.1	146285	9	AC005083	L06178 Apis mellif
7	72	3.1	16343	3	AMFGENOM	AC006970 Homo sapi
8	71.4	3.1	136098	9	AC006970	AC006279 Plasmodiu
9	68.6	3.0	245802	2	AC006279	AL592166 Homo sapi
10	68	2.9	160788	2	AL592166	AL593856 Homo sapi
11	67.8	2.9	156133	2	AL593856	AC072048 Mus muscu
12	67.6	2.9	208684	2	AC072048	AC005506 Plasmodiu
13	66.6	2.9	205429	2	AC005506	AL035477 Plasmodiu
14	66.6	2.9	224448	2	PFMALAP4	AC092304 Homo sapi
15	66.4	2.9	36977	2	AC092304	AE021553 Homo sapi
16	65.2	2.8	185691	2	AC021553	AE021553 Plasmodiu
17	64.4	2.8	14867	3	AE001398	AJ223323 Saccharom
18	64	2.8	1867	8	MTSCAJ23	AL011146 Homo sapi
19	63.8	2.7	158398	2	AC011146	AL158151 Human DNA
20	63.6	2.7	162445	9	AL158151	AX083744 Sequence
21	63.4	2.7	1141	6	AX083744	AC091925 Homo sapi
22	63	2.7	162914	2	AC091925	AL596253 Homo sapi
23	63	2.7	173349	2	AL596253	AL596253 Homo sapi
24	62.8	2.7	173349	2	AL596253	AC068363 Homo sapi
25	62.8	2.7	186529	2	AC068363	AC090493 Homo sapi
26	62.8	2.7	213530	2	AC090493	AC022281 Homo sapi
27	62.6	2.7	186431	2	AC022281	AC021992 Homo sapi
28	62.6	2.7	193302	2	AC021992	AF288092 Naegleria
29	62.4	2.7	49843	3	AF288092	M80833 Pea lox2-1i
30	61.6	2.6	958	8	PEAL21LG	AL445219 Homo sapi
31	61.6	2.6	103649	2	AL445219	AC005969 Homo sapi
32	61.4	2.6	147244	9	AC005969	AL049180 Plasmodiu
33	61.4	2.6	192581	2	PFMALI3P1	AC012627 Homo sapi
34	61.2	2.6	100925	9	AC012627	AC026839 Homo sapi
35	61.2	2.6	156933	2	AC026839	AC021553 Homo sapi
36	61.2	2.6	185691	2	AC021553	AC008739 Homo sapi
37	61.2	2.6	235532	9	AC008739	AL031746 Plasmodiu
38	60.8	2.6	67970	3	PFMALIP3	AP002820 Oryza sat
39	60.8	2.6	137332	8	AP002820	U4117204 Caenorhab
40	60.8	2.6	260341	3	CEY116A8C	U49822 Saccharomyc
41	60.6	2.6	2426	8	SDU49822	M97514 Saccharomyc
42	60.6	2.6	8622	8	YSCMTCYTOC	AC092304 Homo sapi
43	60.6	2.6	36977	2	AC092304	AC073269 Homo sapi
44	60.6	2.6	137889	9	AC073269	AL356793 Homo sapi
45	60.2	2.6	140455	2	AL356793	

ALIGNMENTS

RESULT 1

AB007911

LOCUS

AB007911

DEFINITION

AB007911

ACCESSION

AB007911.1

VERSION

KEYWORDS

SOURCE

ORGANISM

Pism sativum

DNA

3441 bp

PLN

28-SEP-1999

complete cds.

GI:5926717

99292944

2 (bases 1 to 3441)

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

2 (bases 1 to 3441)

1 (sites)

Inaba,T., Nagano,Y., Sakakibara,T. and Sasaki,Y.

Identification of a cis-regulatory element involved in phytochrome

down-regulated expression of the pea small GTPase gene pra2

Plant Physiol. 120 (2), 491-500 (1999)

99292944

2 (bases 1 to 3441)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae;
Pisum.

AUTHORS Nagano,Y.
 TITLE Direct Submission
 JOURNAL Submitted (07-OCT-1997) Yukio Nagano, Nagoya University, Graduate School of Biological Sciences; Chikusa, Nagoya, Aichi 464-8601, Japan [E-mail:nagano@agr.nagoya-u.ac.jp, Tel:81-52-789-4168, Fax:81-52-789-4296]
 FEATURES
 source
 1..3441
 /organism="Pisum sativum"
 /db_xref="taxon:3888"
 repeat_unit
 gene 904..1018
 Join(2326..2567,2729..3188)
 /gene="pra2"
 CDS Join(2326..2567,2729..3188)
 /gene="pra2"
 /note="light-repressible GTP binding protein"
 /codon_start=1
 /product="Pra2"
 /protein_id="BA84640.1"
 /db_xref="GI:5926718"
 /translation="MNQSMNGVEAEKLOEKIDYVFKVVIGDSAVGKTOILSRFTKNE
 FCFDSKSTIGVEFOFKTYTINGKLKAOIWDTAGOERYRAVTSAYIRGALGAMLYVDI
 TKRQFDHVAHWELRSHADGSIVIMLIGNKGLDYDQGVQTEDAVFEADQGLFFS
 ETSFSGENVSASFUKLQEIINKVSKRSLECNKINGNHDVVAALGKEDIILISAS
 ELEISEIKKLHSCSC"
 BASE COUNT 1143 a 606 c 540 g 1152 t
 ORIGIN
 Query Match 100.0%; Score 2325; DB 8; Length 3441;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 aagctttaagcaagggaagacaacaattccaaaataataaaaactcctaaagaatga 60
 DB 1 AGCCTTAAGGCAAGGGAAGACACAAATTCAAAATAATATAAAACTCCTTAAAGAAATGA 60
 QY 61 tttattcttattcctaataaacttttctctatttccaaaaacacacatcaaaagtattgtga 120
 DB 61 TTTTATTCTTATCTTCATAAATAACTTTTCTTATTCAAAACACATCAAAAGTTATGTGA 120
 QY 121 ttcatatttataatctgataataataataattgtatatttcaatttcaataaattgtg 180
 DB 121 TTCATATCTTTAATATCTGATAATAATAATTTGTATATTCATAATATTCATACAATTTG 180
 QY 181 ttatatgaatatatttgttagttaaaggactaaagaataacctccgcaacatcaaaagtca 240
 DB 181 TTATATGAATAATTGTAGTTAAAGGACTAAGAAATACCTCCGCAACATCAAAAGTCA 240
 QY 241 gaaacctctgttaactcttcagttgaaacgagaagaagtggacaacacagaaaaactaag 300
 DB 241 GAAACCTCTGTAACTCTTCAGTTGAAACGAGAAGGAGTGGACAACACAGAAAACTAAG 300
 QY 301 ttcccccaacttaactcttggttgggtgaggaactcccttcaatttatactctaagga 360
 DB 301 TTCCCCCACTTAACCTCTTGGTTGGGTGAGGACTTCCTTTTAACTTATATCTCTAAGGA 360
 QY 361 aaacattagacactctagatgggttgcatagctcatatatttttaagtaataataacc 420
 DB 361 AATACATTAGACACTCTAGATGGGTGCAATAGCTCATATATTTTAAATAATAATACCC 420
 QY 421 acttcaagttttttgtttttttgttggcagtagatgataagatggatcatctttctcaa 480
 DB 421 ACTTCAAGTTTGTGTTTGTGTTGTCGAGTAGATGATAAGATGGATCATTTCTCAA 480
 QY 481 ggcctcttgaaaagacataaagatccataactcccaagaagtgttcttcaatcaacca 540
 DB 481 GGCCCTTATGCAAGACATAAAGATCCATATCTACTCCACCAAGATGCTTTTACATCTAACCA 540
 QY 541 agttaatgaatttaattcttcgaacaataattttctaccagaagaagtatttatatgcac 600
 DB 541 AGTTAATGAATTTAAATCTTCGAAACAATTTATTTCTTACCAAGGAAGTTTATATGACAC 600

QY 601 attttctaattgtattttttatatagaattgatacatgtttctgtttatacaagattagaatt 660
 DB 601 ATTTTCTAATGTATTTTATATAGAATTGATACATGTTCTGTGTATATACAAGATTAGAATT 660
 QY 661 tggatttctatcccaaacctcctacacttgggtgagaaatttccagcctcaacctcagtaaat 720
 DB 661 TGGATTTCATCCCAAACTCCTACACTTGGTGAGAAATTCAGCCCTCAACCTCAGTAAAT 720
 QY 721 caggttctcctcctcaaacctcacaacttgggttgagtgagaatttatgacgtcaaacctagc 780
 DB 721 CAGGTTCTCTCCCTTCAAACTCATACACTTGGTGTAGTGAGAAATATGAGCTCAACCTAGC 780
 QY 781 aatatgaattccctctcccaagatcctacacttatctgagtgagaatttttgctcctcgacct 840
 DB 781 AATATGAATCCCTCTCCAAGATCCTACACTTATCTGAGTGAGAAATTTTGGTCTCTGACCT 840
 QY 841 caacaagatagatttgatgggtcattcacagggggaagacattcacattgggttcaagaattc 900
 DB 841 CAACAAGATAGATTGTGATGGGTCTCACAGGGGGAAGCAATTCACATTGGGTCAAAAGATT 900
 QY 901 acccaaacaaagtgagagagacatcacatatcaacaaaaccttaaggtgatatagggtatg 960
 DB 901 ACCCAAAACAAGTGAGAGACATCATATATCAACCAAAACCTTTAAGGTATAGTGTATG 960
 QY 961 agttctcttacttataaagtgtctcaacctcccacttttctaaagcaatgtgtgacctagaac 1020
 DB 961 AGTTCTCTTACTTATAAAGTGCTCAACCTCCACTTTTCTTAAGCAATGTGTGACTTAGAAC 1020
 QY 1021 tcacacttattctcaacaataactcaacacttgtttatcaacaatctcccccaagaattg 1080
 DB 1021 TCACACTTATTCTCAACATAACTCACACTTGTGTATCAACAATCTCCCCCAACAAGTGTG 1080
 QY 1081 agtcattcgtctatgtccccctcaagtgaattctcttccatccgcatgcttataccgttg 1140
 DB 1081 AGTTCTATTGCTGTATGTCCTCCCTCAAGTGGAAATCTCTTTTATCCGCATGTTATACCGT 1140
 QY 1141 ttgacatacatatttactcgtcatgggcaacttcaatgggacacgctgcctgaccacactg 1200
 DB 1141 TTGACATACATCTTTACTCGTCATGGGCATCTCAATGGGACACGCTGCCTGACCACCATG 1200
 QY 1201 tcaagaagacttttgacacaaaggagtcggtcccttactggaaccagactctgataccatt 1260
 DB 1201 TCAAGAAGACTTTTGACACAAGAGTCGGTCCCTTACTCGAACACACTCTGTATACCAT 1260
 QY 1261 aatagatcactttgaatgatatcattcatactatactatacaaacatttaccgtaagaataaa 1320
 DB 1261 AATAGATCACTTTGAATGGATATCATCTACTATATCAACACTTTACGTAAGATATAAA 1320
 QY 1321 aattcaccacaaacaaatgagagacactacatctctcttattatataataaaaaatgtaa 1380
 DB 1321 AATTCAACCAAAACAAATGAGAGACACTACATCTCTCTTATATATTAATAAAATGTAA 1380
 QY 1381 agaaaaataataataaaaagtaaacacataattttgataaaaatttataaactattttc 1440
 DB 1381 AGAAAAATATAGTATAAAAAGTAACACATATTTTGTATAAATTTATTACTAAAACTATT 1440
 QY 1441 tagtacctgttaactcgtcgtgagattttacagtaataaagaaacgaggtagcccaaac 1500
 DB 1441 TAGTACTTGTTAATCATGTCTGAGGATTTTACAGTAATAAAGNAACGAGGTAGCCCAAC 1500
 QY 1501 aaaagtgaataattgagggtgtgactttgtcgggtgcaaaaaaaatgaaccccccaactt 1560
 DB 1501 AAAAGTGAATAATTGTGGAGGTGTGATCTTTGTGGTGCAGAAAAAATGAACCCCAACTT 1560
 QY 1561 gtgatatgtgtcgactgctccgtcgctacattgaaattaaatgaatgttcttttataacg 1620
 DB 1561 GTGATATTGTGCGACTGCTCCGTCGCTACATTGAAATTAATGAATGTCTTTTATAACG 1620
 QY 1621 ttgtctatgcccgtattaccctatattgctcactagaaatgggacaatgaatttaatatat 1680
 DB 1621 TTTGTCTATGCCGTATTACCCATATGCTCAGTAGAATGGGACANCTGAATTTAATATATAT 1680
 QY 1681 ctgtcatgtgtgggtgatttcaatttaattgtatctgataatggtaggacatactatgct 1740

Db 1681 CTGTCATGTGGTGGATTCAATTTAATGATCGTAAATGGTAGGACATACATCGCT 1740
Qy 1741 acacaattatcatcatcactgctcaatcactgctcaatgctgtttcttctccatgaattc 1800
Db 1741 ACACAAATATATCATCTGCTCAATCAGTGGTCAATGCTGCTCTCTCCCATGAATTC 1800
Qy 1801 acattgtaagaataattaccacaccttaaaatgtttatcccttgccacacatttcacatcaa 1860
Db 1801 ACATTTGCTAAAGAAATATACCACCTTAAATGTTATCCCTTGCACACATTTACATCAA 1860
Qy 1861 ttatttaaacattttaccatttggaacacacacatcatattcaatatttttgatt 1920
Db 1861 TTTATTAAAAATTTTACCATTGGAACACACATACATATTCAATCAATATTATTGCAAT 1920
Qy 1921 ttcaaaactaaacaaacactagaattatttgaattattatagcaaattttcaaaaa 1980
Db 1921 TTCAAAAACCTAAACCAACAACTAGAAATTTTGTAAATTTATAGCAAAATTTTCAAAAA 1980
Qy 1981 tatcctagtcttcacacactcaataattcacaattttccaaatcccttgcaaaaacatcaca 2040
Db 1981 TATCCTAGTCTTCAACCACTCAATAATTCACAATTTCCAAATCCCTTGCAAAAACATCACA 2040
Qy 2041 acctctgaaactttgatttaataatcttaataaaagcaataatgatgatctataaacaatat 2100
Db 2041 ACCCTAGAAACTTTGATTAAATATCTAATAATCTAATAAAGCAATAATATGATATCTAAACAATAT 2100
Qy 2101 caccatatgttatgataataatgatgacgaataacacacttaatttggttaaacattaa 2160
Db 2101 CACCATATATGTTATGATATAATATGATGACGCAATACACTTAATTTGGTAAAGCATTA 2160
Qy 2161 agcgagacaacttataaacacggtaattcaacaacggttgttgctgagttcattgttt 2220
Db 2161 AGCGAGACAACCTATTAAACACCGTAATTCAACACCGTTGTTGTCAGTTTCATGTTT 2220
Qy 2221 cttccaaactcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 2280
Db 2221 CTTCCAACTCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2280
Qy 2281 atatatactctcttgaacctcttcttcttcttcttcttcttcttcttcttcttcttctt 2325
Db 2281 ATATATACTATCTCTCTTGAAACCHCTTTTGTGATCTTGACAAGAAA 2325

RESULT 2
AF138704 20785 bp DNA PLN 03-FEB-2000
LOCUS Pisum sativum gibberellin c20-oxidase gene, complete cds.
DEFINITION AF138704
ACCESSION AF138704
VERSION AF138704.1 GI:6855710
KEYWORDS
SOURCE pea.
ORGANISM Pisum sativum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.
REFERENCE 1 (bases 1 to 20785)
AUTHORS Martin,D.N., Proebsting,W.M., Parks,T.D., Dougherty,W.G., Lange,T., Lewis,M.J., Gaskin,P. and Hedden,P.
TITLE Feed-back regulation of gibberellin biosynthesis and gene expression in Pisum sativum L
JOURNAL Planta 200 (2), 159-166 (1996)
MEDLINE 97060758
REFERENCE 2 (bases 1 to 20785)
AUTHORS Martin,D.N. and Proebsting,W.M.
TITLE Pisum sativum gibberellin c20-oxidase, genomic clone
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 20785)
AUTHORS Martin,D.N. and Proebsting,W.M.
TITLE Direct Submission
JOURNAL Submitted (25-MAR-1999) Horticulture, Oregon State University, ALS 4017, Corvallis, OR 97331-7304, USA

FEATURES
source Location/Qualifiers
1..20785
/organism="Pisum sativum"
/strain="Alaska"
/db_xref="taxon:3888"
join(<1912..2468,2821..3142,3416..3880)
/product="gibberellin c20-oxidase"
join(1912..2468,2821..3142,3416..3679)
/function="converts GA53 to GA44 to GA19 to GA17 and GA20"
/note="oxoglutarate-dependent dioxygenase; similar to the product encoded by GenBank Accession Number X91658"
/codon_start=1
/product="gibberellin c20-oxidase"
/protein_id="AAF29605.1"
/db_xref="GI:6855711"
/translation="MAECITSSAKLMTQKSDKNEGSKLVFDASFLNQLNLPKQ FWPDDKPCMNVPDLIDFNKFLSGDFFFAAMEASKTIGEACERHGFLLVNNHGI DTLLIEHAHSYMNDFFVPLSQKQRCQKGEHCYASFTGRFSSNLPKMETLSQCF SDEKNSNIIVKDYLSNTLGEDFOQGEVYOYCEAMSKLSLIGMELLMGSLGVGKECF RDPFEENKSMRLNYPPCKPDLTGLTGPCDPTSLTILHDOVGGLGVFDNEMHS IRDNFAFVNVIGDTPMALSGRYKSLHRAVNNKTRKSLAFELCPKGDVVSPPS ELYNDLTPRIYPOFTWPMLEFETQKHRYRADMRILEAFTKWIQQQN"
3'UTR 3680..3880
misc_feature 13001..13800
/note="similar to tobacco retroelement presented in GenBank Accession Number X13777"
misc_feature 14272..14582
/note="similar to region upstream of pea chalcone synthase gene presented in GenBank Accession Number X80007"
repeat_region 15844..15905
/rpt_type=tandem
/rpt_unit=at
BASE COUNT 6770 a 3079 c 3474 g 7431 t 31 others
ORIGIN
Query Match 4.3%; Score 100; DB 8; Length 20785;
Best Local Similarity 84.8%; Pred. No. 1.8e-07;
Matches 112; Conservative 0; Mismatches 20; Indels 0; Gaps 0;
Qy 963 ttctcttactataaagtctcaacctccacttttcttaagcaatgtgtgacttagaactc 1022
Db 3 TCCTCTAACTTATAAAGTGTTCACATCCACTTTTCTAAGCAATGTGAGACTTAGAATC 62
Qy 1023 acactttattctcaatacactcaactgttttattcaacaactctccccaagtgaag 1082
Db 63 ACACCTGTGTTCTTACGACAAATTCACATTTGTTTCTTAATATTTCTCCCTGAAAGTGAG 122
Qy 1083 ttcatcgcctat 1094
Db 123 TACATCCACTAT 134
RESULT 3
PSBLOX12
LOCUS PSBLOX12 5785 bp DNA PLN 06-APR-1995
DEFINITION P.sativum (Birt) Loxl:Ps:2 gene.
ACCESSION X78580
VERSION X78580.1 GI:469153
KEYWORDS lipoxxygenase; Loxl gene.
SOURCE pea.
ORGANISM Pisum sativum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.
REFERENCE 1 (bases 1 to 5785)
AUTHORS Casey,R.
TITLE Direct Submission
JOURNAL Submitted (01-APR-1994) R. Casey, John Innes Institute, Colney Lane, Norwich, NR4 7UH, UK
REFERENCE 2 (bases 1 to 5785)

University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

COMMENT

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc/index.shtml>

----- Project Information -----
Center project name: H NH0505B03

* NOTE: This is a 'working draft' sequence. It currently
* consists of 42 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1568: contig of 1568 bp in length
* 1569 1668: gap of unknown length
* 1669 3345: contig of 1677 bp in length
* 3345 3446: gap of unknown length
* 3446 4925: contig of 1480 bp in length
* 4925 5025: gap of unknown length
* 5025 7252: contig of 2227 bp in length
* 7252 7353: gap of unknown length
* 7353 9190: contig of 1838 bp in length
* 9190 9290: gap of unknown length
* 9290 9231: contig of 1728 bp in length
* 9231 11018: gap of unknown length
* 11018 13028: contig of 1910 bp in length
* 13028 13129: gap of unknown length
* 13129 14658: contig of 1530 bp in length
* 14658 14759: gap of unknown length
* 14759 16766: contig of 2008 bp in length
* 16766 16866: gap of unknown length
* 16866 19394: contig of 2528 bp in length
* 19394 19495: gap of unknown length
* 19495 22382: contig of 2888 bp in length
* 22382 22483: gap of unknown length
* 22483 26670: contig of 4188 bp in length
* 26670 26771: gap of unknown length
* 26771 29899: contig of 3129 bp in length
* 29899 30000: gap of unknown length
* 30000 32393: contig of 2394 bp in length
* 32393 32494: gap of unknown length
* 32494 35161: contig of 2668 bp in length
* 35161 35261: gap of unknown length
* 35261 39253: contig of 3992 bp in length
* 39253 39353: gap of unknown length
* 39353 42753: contig of 3400 bp in length
* 42753 42854: gap of unknown length
* 42854 45570: contig of 2717 bp in length
* 45570 45671: gap of unknown length
* 45671 50008: contig of 4338 bp in length
* 50008 50109: gap of unknown length
* 50109 55872: contig of 5764 bp in length
* 55872 55973: gap of unknown length
* 55973 60627: contig of 4655 bp in length
* 60627 60728: gap of unknown length
* 60728 64794: contig of 4067 bp in length
* 64794 64894: gap of unknown length
* 64894 69682: contig of 4788 bp in length
* 69682 69782: gap of unknown length
* 69782 74783: contig of 5000 bp in length
* 74783 74883: gap of unknown length
* 74883 80238: contig of 5356 bp in length
* 80238 80339: gap of unknown length
* 80339 86095: contig of 5757 bp in length
* 86095 86196: gap of unknown length
* 86196 93489: contig of 7294 bp in length
* 93489 93589: gap of unknown length
* 93589 93590 93590: contig of 5923 bp in length

	FEATURES	Location/Qualifiers
	source	1..326924
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/chromosome="UNK"
		/clone="RP11-505B3"
		1..1568
misc_feature		/note="assembly_name:Contig10"
misc_feature		1669..3345
misc_feature		/note="assembly_name:Contig11"
misc_feature		3446..4925
misc_feature		/note="assembly_name:Contig13"
misc_feature		5026..7252
misc_feature		/note="assembly_name:Contig15"
misc_feature		7353..9190
misc_feature		/note="assembly_name:Contig16"
misc_feature		9291..11018
misc_feature		/note="assembly_name:Contig17"
misc_feature		11119..13028
misc_feature		/note="assembly_name:Contig18"
misc_feature		13129..14658
misc_feature		/note="assembly_name:Contig20"
misc_feature		14759..16766
misc_feature		/note="assembly_name:Contig21"
misc_feature		16867..19394
misc_feature		/note="assembly_name:Contig22"
misc_feature		19495..22382
misc_feature		/note="assembly_name:Contig23"
misc_feature		22483..26670
misc_feature		/note="assembly_name:Contig24"
misc_feature		26771..29899
misc_feature		/note="assembly_name:Contig25"
misc_feature		30000..32393
misc_feature		/note="assembly_name:Contig26"
misc_feature		32494..35161
misc_feature		/note="assembly_name:Contig27"
misc_feature		35262..39253
misc_feature		/note="assembly_name:Contig28"
misc_feature		39354..42753
misc_feature		/note="assembly_name:Contig29"
misc_feature		42854..45570
misc_feature		/note="assembly_name:Contig30"
misc_feature		45671..50008
misc_feature		/note="assembly_name:Contig31"
misc_feature		50109..55872

		/note="assembly_name:Contig32" 55973..50627 /note="assembly_name:Contig33" 60728..64794 /note="assembly_name:Contig34" 64895..69682 /note="assembly_name:Contig35" 69783..74782 /note="assembly_name:Contig36" 74883..80238 /note="assembly_name:Contig37" 80339..86095 /note="assembly_name:Contig38" 86196..93489 /note="assembly_name:Contig39" 93590..99512 /note="assembly_name:Contig40" 99613..105968 /note="assembly_name:Contig41" 106069..112079 /note="assembly_name:Contig42" 112180..118180 /note="assembly_name:Contig43" 118281..126420 /note="assembly_name:Contig44" 126521..134015 /note="assembly_name:Contig45" 134116..143412 /note="assembly_name:Contig46" 143513..151892 /note="assembly_name:Contig47" 151993..164752 /note="assembly_name:Contig48" 164853..179948 /note="assembly_name:Contig49" 180049..202305	
		Query Match 3.5%; Score 82.4; DB 2; Length 326924; Best Local Similarity 44.9%; Pred. No. 0.00011; Matches 396; Conservative 0; Mismatches 481; Indels 5; Gaps	
misc_feature		QY 1245 agactctgataccattaataagatcaacttgaatggatcatcattccactacatatcaaacat 1304	
misc_feature		Db 292589 ATAATATATTATATATAAAATATATATTTAATATATATATATATATAAAATATATATATAA 292530	
misc_feature		QY 1305 ttacgtaaagataaaaaattcacccaacaagtgcagagagacactacat-cctctttat 1363	
misc_feature		Db 292529 TATATTATATATAAATATATATATTTTAATATATATATATAAAATATATATTATAAATA 292470	
misc_feature		QY 1364 atattaataaatgtaagaataatagtagtaaaaagtaacacatatcttgtataaaat 1423	
misc_feature		Db 292469 TTATATATAAAATATATATTAAATATATTTATATATATAAAATATATATTATATATATA 292410	
misc_feature		QY 1424 ttactaaacacttttctagtcacttgtaaactgctgaggatttcacagtaataaga 1483	
misc_feature		Db 292409 AAATATATATTTAAATATATATATAAAAATATATATTTTATAATATATATATAAAAA 292350	
misc_feature		QY 1484 aacgaggtagcccaacaagaatgtaattgtggagggtgactcttgcggtgcacaaa 1543	
misc_feature		Db 292349 ATATATTTATATATTTATATATAAAATATATATTTTAAATATATATATTTATAATATAT 292290	
misc_feature		QY 1544 aaatgaacccccaaacttgtgatattgtgtcgactgctccgctcacattgaaatgaatg 1603	
misc_feature		Db 292289 TATATAAAATATATATTATAAATATATATATATAAATATATATAAATATATATATTT 292230	
misc_feature		QY 1604 aatgtctctttataaocqtttgtctatgcgcgatatacccataatggtcactagaatggaca 1663	
misc_feature		Db 292229 AATATATATATGATATATATATATATAAAATATATATTTTAAATATATATATTTATAATA 292170	
misc_feature		QY 1664 atgaatttaatatatctgtcatgtgtgggtgagatcaatttaattgattgcgtaaatgg 1723	
misc_feature		Db 292169 TATTTATATAAAATATATATTTTAAATATATTTATATAAATAATA----TATATTATAATA 292114	

Qy	1724	taggacatactcgtctcacacaatatatcatcatcactcggtccaatccactggccaattggtttt	1783
Ds	292113	TATTATATATAAAAAATATATATTTATTAATATATTTTAAATATAAAAATATATATTAATAATATAT	292054
Qy	1784	tctctcccgaattcacattgctaaagaaaattaccaccttaaaatggttatcccttg	1843
Ds	292053	TATATATAAATATATATTTAATATATTTATATATAAAATATATTTTATAATATATTATA	291994
Qy	1844	cacacatttcacatcaatttatttaaaccattttaccattggaaaaacacacataattcaa	1903
Ds	291993	TATAAATACATATTTATATATAAAAAAATTTATATTTTATTAATATATATATATATA	291934
Qy	1904	tcaattattttgcattttccaaaaactaaacacaaacttagaatatttgttaattat	1963
Ds	291933	TAATTTATATAATATATTTATATAAAAAATATATAAATATATATTTTATAAAAAATATAT	291874
Qy	1964	agcacaattttcaaaaatatctcagtcttcacacacacataaattccaaatttccccaatc	2023
Ds	291873	ATAATATATATTTATATAAAATATATATAATATATATTTGCTATATAAAAAATATAAAATAT	291814
Qy	2024	cttgcaaaaacatcacacacctctagaaccttgatttaataatctcaataaaagcaataata	2083
Ds	291813	ATATAATATATTTGTATATAAAATATATAATATATATTTGCTATATAAAATATATAATA	291754
Qy	2084	tgatatctaaacaatatcaccatatgttatgatataaatat	2125
Ds	291753	TATAATGTATATAAAATATATAAATATATATATTGCTATATAAT	291712
RESULT	5		
PSAI7468			
LOCUS	PSAI7468	1350 bp DNA	PLN 09-JUN-1999
DEFINITION	Pisum sativum Lox2 promoter.		
ACCESSION	Y17468		
VERSION	Y17468.1	GI:5042292	
KEYWORDS	lipoxygenase; lox gene.		
SOURCE	pea.		
ORGANISM	Pisum sativum		
REFERENCE	1 (bases 1 to 1350)		
AUTHORS	Forster,C., North,H., Afzal,N., Domoney,C., Hornostaj,A., Robinson,D.S. and Casey,R.		
TITLE	Molecular analysis of a null mutant for pea (Pisum sativum L.) seed lipoxygenase-2		
JOURNAL	Plant Mol. Biol. 39 (6), 1209-1220 (1999)		
MEDLINE	99308514		
REFERENCE	2 (bases 1 to 1350)		
AUTHORS	Forster,C.		
TITLE	Direct Submission		
JOURNAL	Submitted (05-JUN-1998) C. Forster, John Innes Centre, Norwich Research Park, Norwich, NR4 7UH, UK		
FEATURES	Location/Qualifiers		
source	1..1350		
	/organism="Pisum sativum"		
	/cultivar="Birte"		
	/db_xref="taxon:3888"		
	/note="Near-isogenic LOX-2 null line back-crossed to 'Birte'."		
gene	1..1350		
	/gene="lox"		
promoter	1..1350		
	/gene="lox"		
	/note="Lipoxygenase-2 promoter from LOX-2-null mutant near-isoline"		
BASE COUNT	481 a 238 c 212 g 419 t		
ORIGIN			

	Matches	130;	Conservative	0;	Mismatches	40;	Indels	4;	Gaps	2;
Qy	864	atcacgagggaaagcattccacattgggtc	aaagattc	ccccaaacaagtgagagacat	923					
Db	81	ATCATCAGGAGGACATCCACACTGGT	GC	AAAAAGT---	AACACACAAAGTGAGAGACGT	137				
Qy	924	cacata-tcaacccaaaccttaaggtgat	ggtgat	tgatgtctcttacttataaagtgc	982					
Db	138	CACATATTTACCCAAAATCTTAAGCC	AT	TAGGTAGATGGGTCTCTCACTT	TATAAAGTAA	197				
Qy	983	tcaacctccacttttctaagcaatgtgt	gacttg	aaactcaactattattctca	1036					
Db	198	TCAACCTCCACTTTCTTAAGTAAATG	TGGAACTT	ACAAACCCCTCTCATGTATGA	251					

RESULT	6	
AC005083		
LOCUS	AC005083	DNA
DEFINITION	Homo sapiens BAC clone CTA-281G5 from 7p15-p21, complete sequence.	21-DEC-1999
ACCESSION	AC005083	
VERSION	AC005083.1	GI:4150930
KEYWORDS	HTG.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
TITLE	1 (bases 1 to 146285)	
JOURNAL	Madsen, C. and Blair, T.	
REFERENCE	The sequence of Homo sapiens BAC clone CTA-281G5	
AUTHORS	Unpublished	
TITLE	2 (bases 1 to 146285)	
JOURNAL	Waterston, R. H.	
REFERENCE	Direct Submission	
AUTHORS	Submitted (12-JUN-1998) Genome Sequencing Center, Washington	
TITLE	University School of Medicine, 4444 Forest Park Parkway, St. Louis,	
JOURNAL	MO 63108, USA	
REFERENCE	3 (bases 1 to 146285)	
AUTHORS	Waterston, R.	
TITLE	Direct Submission	
JOURNAL	Submitted (12-JAN-1999) Department of Genetics, Washington	
REFERENCE	University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA	
AUTHORS	4 (bases 1 to 146285)	
TITLE	Waterston, R.	
JOURNAL	Direct Submission	
COMMENT	Submitted (21-DEC-1999) Department of Genetics, Washington	
	University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA	
	On Jan 12, 1999 this sequence version replaced gi:3212908.	
	----- Genome Center	
	Center: Washington University Genome Sequencing Center	
	Center code: WUGSC	
	Web site: http://genome.wustl.edu/gsc	
	Contact: saplens@watson.wustl.edu	
	----- Summary Statistics	
	Center project name: H_RG281G05	

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington

University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GB/CHR7>, send <mailto:regreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
Clone CTA-281C5 is from a release of the human BAC library CTRB-HS-A. The library contains cloned DNA from human sperm. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992). U-1 Kim et al., Genomics 34:213-8 (1996). The clone is available from Research Genetics, Inc. (<http://www.resgen.com>).
VECTOR: pBeloBAC11
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:
Actual start of this clone is at base position 1 of CTA-281G5;
actual end is at 146285 of CTA-281G5

BAC CTA-281G5 contains an *E. coli* transposon from 10982 to 12345 that is not represented in the submitted sequence.

FEATURES	Location/Qualifiers
source	1..146285
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="7"
	/map="7p15-p21"
	/clone="CTA-281G5"
	/clone_lib="CITB-HS-A"
repeat_region	322..358
	/rpt_family=" (TAAAA)n"
repeat_region	808..952
	/rpt_family="MER1_type"
repeat_region	1172..1194
	/rpt_family="AT_rich"
repeat_region	1264..1720
	/rpt_family="L1"
repeat_region	1848..1881
	/rpt_family="AT_rich"
repeat_region	1882..2171
	/rpt_family="Alu"
repeat_region	3782..3804
	/rpt_family="AT_rich"
repeat_region	4794..4826
	/rpt_family="AT_rich"
repeat_region	5764..6034
	/rpt_family="purine-rich"
repeat_region	6891..7126
	/rpt_family="L1"
repeat_region	7133..7263
	/rpt_family="L1"
repeat_region	7455..7481
	/rpt_family="AT_rich"
repeat_region	7504..7554
	/rpt_family="AT_rich"
repeat_region	8320..8348
	/rpt_family="(CA)n"
repeat_region	8466..8533
	/rpt_family="AT_rich"
repeat_region	9611..9733
	/rpt_family="MTR"
repeat_region	9946..10172
	/rpt_family="MER1_type"
repeat_region	10437..10477
	/rpt_family="(CA)n"
repeat_region	10581..10681
	/rpt_family="(TAAA)n"
repeat_region	10759..11049
	/rpt_family="AT_rich"
repeat_region	11050..11235
	/rpt_family="(CAATA)n"
repeat_region	11408..11526
	/rpt_family="(CATAT)n"
repeat_region	11768..11872

[illegible]

tRNA	<pre>/note="codon recognized: AAG" /product="tRNA-Arg" /anticodon=(pos:6602..6604,aa:Arg) 6734..6802 /note="codon recognized: AAC" /product="tRNA-Asn" /anticodon=(pos:6765..6767,aa:Asn) complement(6810..6878) /note="codon recognized: AUU; anticodon unspecified" /product="tRNA-OTHER" complement(6892..8556) /codon_start=1 /transl_table=5 /product="NADH dehydrogenase subunit 5" /protein_id="AAB96805.1" /db_xref="GI:829007" /translation="MIKMWCGGILLPEFSLMMLSLYLLYLNKEFFFEWNIYFNSM KFNELLIIDYKSLMFIPLVSMIFYSIMYDSELMKDRFLYMLILFLISMYTL ILSPNMLGISILGWDGLGISVCLIVYIYMKMSKPTSGMTVILLNRLGDIGLLIMGLT YGSWNLSFYKNNEFNMFIYLLMAFTSAQIIPSTWLPMAAMAPTVPVSVLHVSHSTV AGYILLIYNNIDFNKNTMLTASITMLFAGLVANFELDLKKVAYGYIYIPMKSMI SMLSGISTELVFLHLFIHAFKSLMFCVGSYMHYISNODIRMYGYMIYIPKSMI LIFSILSCGPPFIVGYSKDILIEEMFFSKMIYFSMNLIIGTITFTVSYSFMILV TSKFLMNNYIYSKEDTKICISMMMMIIFSLIYKLIPLMNFNLGLILMILYKLVAF KIMVGLIGMPNFYKILNANKIKYFKMSLEFNLIYKIYKKIIMMMTEYVEYIEKS IIEILSKSMVSTVNIYELKISNLMINIYITILYIYLIYLIYLINE" complement(8557..8624) /note="codon recognized: CCA" /product="tRNA-His" /anticodon=(pos:8589..8591,aa:His) complement(8644..9987) /codon_start=1 /transl_table=5 /product="NADH dehydrogenase subunit 4" /protein_id="AAB96806.1" /db_xref="GI:552446" /translation="MYLLLLILMLMLMMSMIYLFMLFMKNKNNLNLIIGNLIILN LLNFNLNIDWIYIFCNLSFNMYSYGLIMLTWIFGLIFISLNNSLNCGLFNLLIM LSIYLVFLSNLLLFYIFEGGLLIYFLVVKVGSYENRWLSGFYLMFYTMIFSLMP YIYIYLYIDSLNFMELMNLNLNMLFYILMSPFLKIPYIYLPFGWMLLKRAVAP YGSWILASIMLKUGGYKMLIYIYNEFTIQLIVLINSFGVLISLUMCLISQDFM KSITASISIVHMLGMWMTFLKISIGYLLMWSHGLSSGCLFVLNVIYSQTSNR LMFNKGMINPMPKSNLLPWLCSNNGSPVSNLISEVMLLIGMISWAKFMMILIM YCLFSYISYILYFPMFINHGKIFIMFKINGLILVEYFVLLLHPIPLNLMFUKLYFI" complement(9991..10254) /codon_start=1 /transl_table=5</pre>
tRNA	
CDS	
CDS	
CDS	

	Query Match	3.1%	Score 72;	DB 3;	Length 16343;
	Best Local Similarity	44.8%;	Pred. No.	0.008;	
Matches	361; Conservative	0;	Mismatches	440;	Gaps 5;
QY	1317	aaaaaatcccccacaaatgagagagacactcacattcttcttatataattaataaaaat	1376		
Dd	1548	AATCAATTAAACAAATAAACAACACTAACACATGATTATTTAATAAATTTATAAT	15517		
QY	1377	gtaaagaaaaatatgataaaagttaaacacaatttttgataaaatttattaccataaaaacct	1436		
Dd	15518	AAATAATATAAAATAATTTTTTAAATATATATATATATATATATATATATATATTAATTT	15577		
QY	1437	ttttcagacttgtgaatcatgcctggaggatttttacagtaaagaacagvgtagcccc	1496		
		TATTTTTCATAGTANTTAATATATAATTTTATTTTAGTATAAAAATTTAAATATAA	15637		
Dd	15578				
QY	1497	aaccaaagtgataattcgtggagggtgatctcttgcggtcgcaaaaaatgaacccccaa	15556		
Dd	15638	AATTAACATATTTTATATATAAAAATATTTTTTTTTTATTAATTAATTTAAATATAAATA	15697		
QY	1557	acttgatatttgctgcgactgccctgcgcatcatactaataatgaagtctctttat	1616		
Dd	15698	ANTAATAAANGAGCTTTTTTTTTTTTTTGTTATTTTATTATTTTAAATAAANAATATA	15757		
QY	1617	aacggttgctctatgccgtattaccocatatgctocactagaalggagcaaatgaatttaat	1676		

[illegible][illegible]

AL593856.6	GI:15041959	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.	human.	Homo sapiens	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	1 (bases 1 to 156133)	McLay, K.	Direct Submission	Submitted (28-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk	On Jul 30, 2001 this sequence version replaced gi:15028764.	----- Genome Center Center: Sanger Centre Center code: SC Web site: http://www.sanger.ac.uk Contact: humquery@sanger.ac.uk ----- Project Information Center project name: bA8J9 ----- Summary Statistics Assembly program: XGAP4; version 4.5 Sequencing vector: plasmid; L08752; 100% of reads Chemistry: Dye-terminator Big Dye; 100% of reads Consensus quality: 153103 bases at least Q40 Consensus quality: 153756 bases at least Q30 Consensus quality: 154302 bases at least Q20 Insert size: 155533; sum-of-contigs Quality coverage: 163653; 1.8% error; agarose-fp Quality coverage: 6.73x in Q20 bases; sum-of-contigs Quality coverage: 6.48x in Q20 bases; agarose-fp ----- * NOTE: This is a 'working draft' sequence. It currently * consists of 7 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * * 1 3197: contig of 3197 bp in length * 3198 3297: gap of 100 bp * 3298 24805: contig of 21508 bp in length * 24806 24905: gap of 100 bp * 24906 27453: contig of 2548 bp in length * 27454 27553: gap of 100 bp * 27554 64419: contig of 36866 bp in length * 64420 64519: gap of 100 bp * 64520 84883: contig of 20364 bp in length * 84884 84983: gap of 100 bp * 84984 133876: contig of 48893 bp in length * 133877 133976: gap of 100 bp * 133977 156133: contig of 22157 bp in length.	
VERSION	AL593856.6	GI:15041959	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.	human.	Homo sapiens	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	1 (bases 1 to 156133)	McLay, K.	Direct Submission	Submitted (28-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk	On Jul 30, 2001 this sequence version replaced gi:15028764.	----- Genome Center Center: Sanger Centre Center code: SC Web site: http://www.sanger.ac.uk Contact: humquery@sanger.ac.uk ----- Project Information Center project name: bA8J9 ----- Summary Statistics Assembly program: XGAP4; version 4.5 Sequencing vector: plasmid; L08752; 100% of reads Chemistry: Dye-terminator Big Dye; 100% of reads Consensus quality: 153103 bases at least Q40 Consensus quality: 153756 bases at least Q30 Consensus quality: 154302 bases at least Q20 Insert size: 155533; sum-of-contigs Quality coverage: 163653; 1.8% error; agarose-fp Quality coverage: 6.73x in Q20 bases; sum-of-contigs Quality coverage: 6.48x in Q20 bases; agarose-fp ----- * NOTE: This is a 'working draft' sequence. It currently * consists of 7 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * * 1 3197: contig of 3197 bp in length * 3198 3297: gap of 100 bp * 3298 24805: contig of 21508 bp in length * 24806 24905: gap of 100 bp * 24906 27453: contig of 2548 bp in length * 27454 27553: gap of 100 bp * 27554 64419: contig of 36866 bp in length * 64420 64519: gap of 100 bp * 64520 84883: contig of 20364 bp in length * 84884 84983: gap of 100 bp * 84984 133876: contig of 48893 bp in length * 133877 133976: gap of 100 bp * 133977 156133: contig of 22157 bp in length.
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.	human.	Homo sapiens	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	1 (bases 1 to 156133)	McLay, K.	Direct Submission	Submitted (28-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk	On Jul 30, 2001 this sequence version replaced gi:15028764.	----- Genome Center Center: Sanger Centre Center code: SC Web site: http://www.sanger.ac.uk Contact: humquery@sanger.ac.uk ----- Project Information Center project name: bA8J9 ----- Summary Statistics Assembly program: XGAP4; version 4.5 Sequencing vector: plasmid; L08752; 100% of reads Chemistry: Dye-terminator Big Dye; 100% of reads Consensus quality: 153103 bases at least Q40 Consensus quality: 153756 bases at least Q30 Consensus quality: 154302 bases at least Q20 Insert size: 155533; sum-of-contigs Quality coverage: 163653; 1.8% error; agarose-fp Quality coverage: 6.73x in Q20 bases; sum-of-contigs Quality coverage: 6.48x in Q20 bases; agarose-fp ----- * NOTE: This is a 'working draft' sequence. It currently * consists of 7 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * * 1 3197: contig of 3197 bp in length * 3198 3297: gap of 100 bp * 3298 24805: contig of 21508 bp in length * 24806 24905: gap of 100 bp * 24906 27453: contig of 2548 bp in length * 27454 27553: gap of 100 bp * 27554 64419: contig of 36866 bp in length * 64420 64519: gap of 100 bp * 64520 84883: contig of 20364 bp in length * 84884 84983: gap of 100 bp * 84984 133876: contig of 48893 bp in length * 133877 133976: gap of 100 bp * 133977 156133: contig of 22157 bp in length.		
ORGANISM	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.	human.	Homo sapiens	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	1 (bases 1 to 156133)	McLay, K.	Direct Submission	Submitted (28-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk	On Jul 30, 2001 this sequence version replaced gi:15028764.	----- Genome Center Center: Sanger Centre Center code: SC Web site: http://www.sanger.ac.uk Contact: humquery@sanger.ac.uk ----- Project Information Center project name: bA8J9 ----- Summary Statistics Assembly program: XGAP4; version 4.5 Sequencing vector: plasmid; L08752; 100% of reads Chemistry: Dye-terminator Big Dye; 100% of reads Consensus quality: 153103 bases at least Q40 Consensus quality: 153756 bases at least Q30 Consensus quality: 154302 bases at least Q20 Insert size: 155533; sum-of-contigs Quality coverage: 163653; 1.8% error; agarose-fp Quality coverage: 6.73x in Q20 bases; sum-of-contigs Quality coverage: 6.48x in Q20 bases; agarose-fp ----- * NOTE: This is a 'working draft' sequence. It currently * consists of 7 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * * 1 3197: contig of 3197 bp in length * 3198 3297: gap of 100 bp * 3298 24805: contig of 21508 bp in length * 24806 24905: gap of 100 bp * 24906 27453: contig of 2548 bp in length * 27454 27553: gap of 100 bp * 27554 64419: contig of 36866 bp in length * 64420 64519: gap of 100 bp * 64520 84883: contig of 20364 bp in length * 84884 84983: gap of 100 bp * 84984 133876: contig of 48893 bp in length * 133877 133976: gap of 100 bp * 133977 156133: contig of 22157 bp in length.		
REFERENCE	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.	human.	Homo sapiens	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	1 (bases 1 to 156133)	McLay, K.	Direct Submission	Submitted (28-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk	On Jul 30, 2001 this sequence version replaced gi:15028764.	----- Genome Center Center: Sanger Centre Center code: SC Web site: http://www.sanger.ac.uk Contact: humquery@sanger.ac.uk ----- Project Information Center project name: bA8J9 ----- Summary Statistics Assembly program: XGAP4; version 4.5 Sequencing vector: plasmid; L0875		

```
/note="assembly_fragment:03551
fragment_chain:1"
84984..133876
/note="assembly_fragment:01981
fragment_chain:1"
133977..156133
/note="assembly_fragment:01742
clone_end:t7
vector_side:right"
BASE COUNT 46148 a 32182 c 32348 g 44855 t 600 others
ORIGIN
Query Match 2.9%; Score 67.8; DB 2; Length 156133;
Best Local Similarity 41.6%; Pred. No. 0.032;
Matches 395; Conservative 0; Mismatches 542; Indels 12; Gaps 5;
QY 1258 attaatagatcaacttggaatggatatactcattcactatcatcaacaacatttcagcgaagata 1317
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 27841 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 27782
QY 1318 aaaaattccaccacaaatgagagagacactacatctctcttatttatataataaaatg 1377
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 27781 TAAATTTATATATATATATATATATATATATATATATATATATATATATATATATATAT 27722
QY 1378 taagaaaaatagtagtaaaagtaacacacatttttgataaaattttactactaaacattt 1437
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 27721 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 27662
QY 1438 ttctagtaactgttaactatgctgaggattttacagtaataaagaacagaggtagccca 1497
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 27661 TTATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 27603
QY 1498 acaaaagtgataattgtgaggggtgatcttctgctgcaaaaatgaaaccccaaa 1557
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 27602 ATAAAAATATATATATATATATATATATATATATATATATATATATATATATATATATAT 27543
QY 1558 cttgtgatattgtgcactgctccgcgcacatigaaaattaatgaatttctttata 1617
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 27542 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 27483
QY 1618 acgttgtctatccgtattaccataggtcactagaatggagacaatgaatttaata 1677
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 27482 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 27423
QY 1678 tatctgtc-----atgtgtgggtggatttcaatttaattgtatcgtaaatgttaggacata 1732
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 27422 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 27363
QY 1733 ctcatgtcacacaattatcatcactggtgcactcactggtcactggttctctctcc 1792
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 27362 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 27303
QY 1793 atgaat---tcacattgctaagaataattaccaccttaaatgtttatcccttgacaca 1849
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 27302 AAAATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 27243
QY 1850 ttccacatcaattttataaacattttaccattggaaacacacacatatttcaatcaatt 1909
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 27242 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 27183
QY 1910 atttttgcatt-ttcaaaataacacaaacaaacttgagaattttgttaattatagca 1967
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 27182 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 27123
QY 1968 caattttcaaaaatactcctagcttccaccactcaataatttccca-aatccct 2026
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 27122 AAAATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 27063
QY 2027 tgcaaaacatcacaaacctctagaaactttgattaataactaataaaagcaataatga 2086
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 27062 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 27003
```

```
QY 2087 tatctaaacaatatcaccatatatgttatgatataatgatgcagcaatacacacttaatt 2146
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 27002 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 26943
QY 2147 tggtaaaagcaattaaagcgagacaactcttattaaacacgggtaattcaaca 2195
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 26942 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 26894

RESULT 12
AC072048/c 208684 bp DNA HTG 07-JUN-2000
LOCUS Mus musculus chromosome 6 clone RP23-189E15 strain C57BL6/J,
DEFINITION WORKING DRAFT SEQUENCE, 24 unordered pieces.
ACCESSION AC072048
VERSION 1 GI:8313202
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 208684)
AUTHORS Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
Bouffard,G.G., Dietrich,N.L., Eagle,W.O., Gupta,J., Ho,S.-L.,
Huang,M.C., Idol,J., Lee-Lin,S.-Q., Maduro,Q.L., Maduro,V.B.,
Mastrian,S.D., McCloskey,J.C., Morse,E., Ojodu,M.A., Pearson,R.,
Stantripop,S., Summers,T.J., Thomas,J.W., Thomas,P.J.,
Tongson,E.E., Touchman,J.W., Tran,J.T., Vogt,J.L., Walker,M.A.,
Wetherby,K.D. and Green,E.D.
TITLE NISC Mouse Sequencing Initiative
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 208684)
AUTHORS Green,E.D.
TITLE Direct Submission
JOURNAL Submitted (07-JUN-2000) NIH Intramural Sequencing Center, 8717
COMMENT Grovmont Circle, Gaithersburg, MD 20877, USA
-----
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.mouse@hgrl.nih.gov
-----
Project Information
Center project name: WP
Center clone name: 189E15
-----
Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 195465 bases at least Q40
Consensus quality: 197294 bases at least Q30
Consensus quality: 198446 bases at least Q20
Insert size: 209000; agarose-fp
Insert size: 197000; pulse-field-gel
Insert size: 206384; sum-of-contigs
Quality coverage: 6.43x in Q20 bases; agarose-fp
Quality coverage: 6.82x in Q20 bases; pulse-field-gel
Quality coverage: 6.51x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2209: contig of 2209 bp in length
* 2210 2309: gap of unknown length
* 2310 4736: contig of 2427 bp in length
* 4737 4836: gap of unknown length
* 4837 7023: contig of 2187 bp in length
* 7024 7123: gap of unknown length
* 7124 9505: contig of 2382 bp in length
* *
```



```
AC005506      205429 bp      DNA      12-AUG-2000
LOCUS          Plasmodium falciparum chromosome 12 clone 3D7, *** SEQUENCING IN
DEFINITION    PROGRESS ***, 4 unordered pieces.
ACCESSION     AC005506
VERSION       AC005506.8 GI:9797717
KEYWORDS      HTG: HTGS_PHASE1.
SOURCE        malaria parasite P. falciparum.
ORGANISM      Plasmodium falciparum
REFERENCE     1 (bases 1 to 205429)
AUTHORS       Hyman,R.W., Fung,E.L., Qin,F., Rowley,D., Mao,J., Tamaki,T.,
              Kurdi,O.B., Conway,A.B. and Davis,R.W.
TITLE         Plasmodium falciparum 3D7 chromosome 12
REFERENCE     2 (bases 1 to 205429)
AUTHORS       Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.
TITLE         Direct Submission
JOURNAL       Submitted (21-AUG-1998) Stanford DNA Sequencing and Technology
              Center, Stanford University, 855 California Avenue, Palo Alto, CA
              94304, USA
COMMENT       On Aug 12, 2000 this sequence version replaced gi:8810449.
              * NOTE: This is a 'working draft' sequence. It currently
              * consists of 4 contigs. The true order of the pieces
              * is not known and their order in this sequence record is
              * arbitrary. Gaps between the contigs are represented as
              * runs of N, but the exact sizes of the gaps are unknown.
              * This record will be updated with the finished sequence
              * as soon as it is available and the accession number will
              * be preserved.
              *
              * 1 132269: contig of 132269 bp in length
              * 132270 132469: gap of unknown length
              * 132470 192227: contig of 59758 bp in length
              * 192228 192427: gap of unknown length
              * 192428 203864: contig of 11437 bp in length
              * 203865 204064: gap of unknown length
              * 204065 205429: contig of 1365 bp in length.
              *
              * Location/Qualifiers
              *   .205429
              *   /organism="Plasmodium falciparum"
              *   /db_xref="taxon:5833"
              *   /chromosome="12"
              *   /clone="PFYAC357"
              *   /clone="3D7"

BASE COUNT    83932 a 17937 c 19766 g 83194 t      600 others
ORIGIN

Query Match    2.9%; Score 66.6; DB 2; Length 205429;
Best Local Similarity 43.4%; Pred. No. 0.049;
Matches 356; Conservative 0; Mismatches 464; Indels 1; Gaps 1;

QY 1350 acatctcttattataataaagttaagaaataatagataaaagtaaacacata 1409
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 83502 AATCAGATATGTCATTAAATAATAATATATATATATA-TATATATATTATGTT 83560
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1410 ttgtgaaattattactaaaactatttcttagtactgttaactcatgctgagattt 1469
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 83561 TTTTATTATTATTCCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 83620
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1470 tacagtaataagaacagggtagcccaacaaagtgataattgtggggggtgatct 1529
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 83621 CTATTTAGGAAATAAATATATATAGAATAAATAAATAAATAAATAAATAAATAA 83680
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1530 ttgtcggtagcaaaatgaaccccaactgtgatattgtgcgactgctcgcta 1589
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 83681 ATATTAAGTATATAAAGAACATCTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 83740
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1590 cattgaaattaagtgttcttttataacgtttgtctatgcggtattaccocatggtc 1649
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 83741 AATTAATAATTAAGAAATATAAAATTTATGACATATATATATATATATATATATA 83800
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1650 actagaatgggacaatgaatttaataatatatctgtcatgtgtgtgggttcattta 1709
```

```
Db 83801 TATATATATATATATATATATATATATATATATATATATATATATATATATAT 83860
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1710 tgcataagtaagtaggacatactactgcacacattatcatcactggcgaatcac 1769
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 83861 ACAGGTATAAATATATAATATATATACATATATATATATATATATATATATATATA 83920
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1770 tggccaatgtgtttctcttcccatgaattcacattgtcgaagaaattaccacctaaa 1829
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 83921 TTATAATTAAATAAGCATACTATAATATATATTTCTTTTATATCCCAATAAAAA 83980
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1830 atgtttatccctgcacacatttcacatcaattatttaaacaattttaccattggaac 1889
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 83981 ATATTTATGCATATATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 84040
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1890 acatacatattcaatcaattatttttgcattttcacaacaaactaaacaaacaaactagaa 1949
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 84041 TTCAAAATTTTATTTATATGCTATTCATAAAAGTATACATGTTGAAATACACACAAA 84100
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1950 tatttgaattatagcacaaattttcaaaaataatccctagcttcaaccactcaaatc 2009
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 84101 AAAAAAGGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAT 84160
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2010 acaatttccaaatccctgcacacacacacacacacacacacacacacacacacacac 2069
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 84161 ATAAAAATGAAAAATTTAATAAATTCATGATATATATATATATATATATATATATACAAT 84220
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2070 taaaagcaataatgatatactcaacaaatcaccatattgttattgatatataatgatg 2129
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 84221 ATTCAATAATATATATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 84280
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2130 cagcaatacacacttaatttggtaagcattaaagcgagacaa 2170
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 84281 TTCTTATAAGCAAAATTTGAGTAAATATATATAAATAAGGAAA 84321
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 14
PFMAL4P4      224448 bp      DNA      HTG
LOCUS          Plasmodium falciparum chromosome 4 strain 3D7, *** SEQUENCING IN
DEFINITION    PROGRESS ***, in unordered pieces.
ACCESSION     AL035477
VERSION       AL035477.5 GI:5731932
KEYWORDS      HTG: HTGS_PHASE1.
SOURCE        malaria parasite P. falciparum.
ORGANISM      Plasmodium falciparum
REFERENCE     1 (bases 1 to 224448)
AUTHORS       Bowman,S., Churcher,C., Harris,B., Harris,D., Lawson,D., Quail,M.
              and Barrell,B.
TITLE         Direct Submission
JOURNAL       Submitted (24-FEB-1999) P.falciparum Genome Sequencing Consortium,
              The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
              CB10 1SA, UK
COMMENT       On Aug 12, 1999 this sequence version replaced gi:5531400.
              For more information about this sequence or the Malaria Project,
              see http://www.sanger.ac.uk/Projects/P.falciparum. IMPORTANT: This
              sequence is unfinished and does not necessarily represent the
              correct sequence. Work on the sequence is in progress and the
              release of this data is based on the understanding that the
              sequence may change as work continues. The sequence may be
              contaminated with foreign sequence from E.coli, yeast, vector,
              phage etc.
              Order of segments is not known; 800 n's separate segments.
              * NOTE: This is a 'working draft' sequence.
              * This record will be updated with the finished sequence
              * as soon as it is available and the accession number will
              * be preserved.
              *
              * Location/Qualifiers
              *   .224448
              *   /organism="Plasmodium falciparum"
              *   /strain="3D7"
              *   /db_xref="taxon:5833"
```

BASE COUNT 80792 a 23258 c 21957 g 75240 t 23201 others
ORIGIN

/chromosome="4"

Query Match 2.98; Score 66.6; DB 2: Length 224448;

Best Local Similarity 20.1%; Pred. No. 0.049;

Matches 282; Conservative 0; Mismatches 1114; Indels 7; Gaps 2;

```
Qy 36 aaataaaaaactctaaagaatattttattcttctctcctaataaacttttcttactt 95
||||| || ||||| ||||| ||||| ||||| |||||
Db 72021 AAATATTACCTTCATCTATATTATCTACCATATATCTATGAAATAAATACATATA 72080
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 96 ccaaaaaacacataaagtattgtattcattcttcttattcttctgataataataattgt 155
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 72081 CGATAGATATATATTTTAGTATTTTATCATATTGAAATAGTATATTATAAATATA 72140
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 156 atattcaaatcttcaacaatttggttatgaaatattttgtaggtaaaagggaactaa 215
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 72141 ATTTTATTGTTATATAATATAGTTTATAATATATTACAATATATAATGTATTATATAG 72200
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 216 aataacctccgaacatcaaaagtcagaacacctcttgtaacctcttcagttgaaacgaga 275
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 72201 ATATATGTACACATCAAAAAAATAAATAATAT -ACAATACAATAACATAGAGAA 72259
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 276 gaagtggacaacacagaaaactaaagttcccccacttactcttctgttggtggaggact 335
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 72260 GGACTATTGTATTATAAAAAAATACACTTTTATTATATCTATTTATATACTTCA 72319
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 336 tcccttcaaatattactctaaggaaatcacattagacactctagatgggtgcattagct 395
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 72320 TTTTATAAATTTATATATTTTATGTATTCATGTATCTTATATATATCTCTTTATTG 72379
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 396 catatttttaagtaataaccacactcaagttttttttttttttttttgtgtgtgcagta 455
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 72380 C-----TTTATAATATAAAGACATTAAGTGACAAATTTATATATGTATANNNNNN 72433
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 456 gatgataagatggatcattctcaggcccttatgcaagacataagaatccataactcc 515
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 72434 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 72493
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 516 accaagatgtcttacatacctaaccagtaagtgaattaaattctctgaaacaattatt 575
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 72494 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 72553
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 576 cctaccaagaagtttatatgcacatttctaatgtatttttatataagaattgatacat 635
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 72554 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 72613
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 636 gttctgttatacaagattagaatttgattctctcacaactcctacacttggtgaga 695
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 72614 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 72673
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 696 aatttcagcctcaacctcagtaaatcagggtctctctctcaactcatacacttggtgag 755
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 72674 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 72733
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 756 tgagaattatggacgtcaacctagcaatatgaatccctctcaagatcctacattatct 815
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 72734 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 72793
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 816 gagtgaattttggtctctcgacctcaacaagatagattgtgaggtcatcagagggga 875
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 72794 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 72853
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 876 agcattcatattgggtcaaaagattcaccccaacaagtgtagagagacatacaacc 935
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 72854 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 72913
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 936 aaaaccttaaggtgataggtgatgagttcttcttactataaagtgctcaacctcactt 995
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 72914 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 72973
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Qy 996 ttctaagcaatgttgacttagaactcacacttatttctcaacataactcacacttggtt 1055
Db 72974 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 73033
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1056 atcaacaatctccccacaagtgtagtccattcgctatgtccccctcaagtgaatctc 1115
Db 73034 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 73093
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1116 ttctcccgcatgcttataccgttgtagacacacttcttactcgtcagggcaactcaa 1175
Db 73094 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 73153
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1176 tgggacagctgctgaccaccatgtcaagaagacttttgacaaaggagtcggtccctt 1235
Db 73154 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 73213
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1236 actcgaaccagactgataccattaatagatcaccttgaatggatatactcactactat 1295
Db 73214 NNNNNNNNNNNNNNAACAAGTCTATAATAATATATATTTGCTTCTTTAAATTAAT 73273
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1296 atcaaacatttacgtaaaagataaaaaattccaccacaataatgagagacacacatct 1355
Db 73274 TAAAAATATTATCATATTATAAAAAATAAACTTAAACAGCATTTATAAAGAAAAAT 73333
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1356 ctcttattattataaagttaaagaaaatatagtaaaagttaacacatatatttga 1415
Db 73334 TCATGTATATTTTATTAAATAAAAAAATAAAAAAATAAAAAAAGGAACCATATA 73393
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1416 taaatttattactaaaactattt 1438
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 73394 TATATATATATATATATATATAT 73416
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 15

AC092304/c

LOCUS

DEFINITION

AC092304 36977 bp DNA HTG 03-JUL-2001

SEQUENCE, 6 unordered pieces.

AC092304

AC092304.1 GI:14589493

VERSION

HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

KEYWORDS

human.

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 36977)

DOE Joint Genome Institute.

TITLE

Sequencing of Human Chromosome 19

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 36977)

DOE Joint Genome Institute.

AUTHORS

Direct Submission

TITLE

Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

-----Genome Center

Center: Joint Genome Institute

Center Code: JGI

Web site: <http://www.jgi.doe.gov>

Project Information

Center Project Name: 31883, FOS36853

Center clone name: LLNL-FOS_20C5

Summary Statistics

Consensus quality: 31111 bases at least Q40

Consensus quality: 33034 bases at least Q30

Consensus quality: 33871 bases at least Q20

Estimated insert size: 42980; agarose-fp estimation

Estimated insert size: 36477; sum-of-contigs estimation

Quality coverage: 6.56 in Q20 bases; agarose-fp estimation

Quality coverage: 7.73 in Q20 bases; sum-of-contigs estimation.

* NOTE: This is a 'working draft' sequence. It currently

* consists of 6 contigs. The true order of the pieces

* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

	1580: contig of 1580 bp in length
1581	1680: gap of unknown length
1681	3748: contig of 2068 bp in length
3749	3848: gap of unknown length
3849	6070: contig of 2222 bp in length
6071	6170: gap of unknown length
6171	13635: contig of 7465 bp in length
13636	13735: gap of unknown length
13736	22951: contig of 9216 bp in length
22952	23051: gap of unknown length
23052	36977: contig of 13926 bp in length.

Location/Qualifiers

FEATURES	SOURCE
1. <i>General</i>	
2. <i>Specific</i>	
3. <i>Other</i>	
4. <i>Other</i>	
5. <i>Other</i>	
6. <i>Other</i>	
7. <i>Other</i>	
8. <i>Other</i>	
9. <i>Other</i>	
10. <i>Other</i>	
11. <i>Other</i>	
12. <i>Other</i>	
13. <i>Other</i>	
14. <i>Other</i>	
15. <i>Other</i>	
16. <i>Other</i>	
17. <i>Other</i>	
18. <i>Other</i>	
19. <i>Other</i>	
20. <i>Other</i>	
21. <i>Other</i>	
22. <i>Other</i>	
23. <i>Other</i>	
24. <i>Other</i>	
25. <i>Other</i>	
26. <i>Other</i>	
27. <i>Other</i>	
28. <i>Other</i>	
29. <i>Other</i>	
30. <i>Other</i>	
31. <i>Other</i>	
32. <i>Other</i>	
33. <i>Other</i>	
34. <i>Other</i>	
35. <i>Other</i>	
36. <i>Other</i>	
37. <i>Other</i>	
38. <i>Other</i>	
39. <i>Other</i>	
40. <i>Other</i>	
41. <i>Other</i>	
42. <i>Other</i>	
43. <i>Other</i>	
44. <i>Other</i>	
45. <i>Other</i>	
46. <i>Other</i>	
47. <i>Other</i>	
48. <i>Other</i>	
49. <i>Other</i>	
50. <i>Other</i>	
51. <i>Other</i>	
52. <i>Other</i>	
53. <i>Other</i>	
54. <i>Other</i>	
55. <i>Other</i>	
56. <i>Other</i>	
57. <i>Other</i>	
58. <i>Other</i>	
59. <i>Other</i>	
60. <i>Other</i>	
61. <i>Other</i>	
62. <i>Other</i>	
63. <i>Other</i>	
64. <i>Other</i>	
65. <i>Other</i>	
66. <i>Other</i>	
67. <i>Other</i>	
68. <i>Other</i>	
69. <i>Other</i>	
70. <i>Other</i>	
71. <i>Other</i>	
72. <i>Other</i>	
73. <i>Other</i>	
74. <i>Other</i>	
75. <i>Other</i>	
76. <i>Other</i>	
77. <i>Other</i>	
78. <i>Other</i>	
79. <i>Other</i>	
80. <i>Other</i>	
81. <i>Other</i>	
82. <i>Other</i>	
83. <i>Other</i>	
84. <i>Other</i>	
85. <i>Other</i>	
86. <i>Other</i>	
87. <i>Other</i>	
88. <i>Other</i>	
89. <i>Other</i>	
90. <i>Other</i>	
91. <i>Other</i>	
92. <i>Other</i>	
93. <i>Other</i>	
94. <i>Other</i>	
95. <i>Other</i>	
96. <i>Other</i>	
97. <i>Other</i>	
98. <i>Other</i>	
99. <i>Other</i>	
100. <i>Other</i>	

```

location/Qualifiers
1. .36977
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"

```

/clone="LLNLFOS-20C5"			
BASE COUNT	11122 a	7407 c	7254 g 10688 t
ORIGIN			506 others

Query Match 2.9%; Score 66.4; DB 2; Length 36977;
Best Local Similarity 43.8%; Pred. NO. 0.063;
Matches 448; Conservative 0; Mismatches 561; Indels 15; Gaps 3;

QY	1293	tatacaaacattttagcgaagatacaaaaaattccaccaaacaaatgagagagacactaca	1352
DB	33981	TATGTTATATAAATTTATATATATAATAATAATTTATATGTTATATAAATTTATATATA	33922
QY	1353	tctctctattatataataaaaatgtgaagaaaaatagatataaaaagtaaacacatattt	1412
DB	33921	TATATAATAAATTTTATATATATATACACTTACATATATATATATTAATATATAATTAT	33862
QY	1413	tgataaatttattactaaaaactatttctagctactgtttaatcatgctcgaggattttac	1472
DB	33861	ATATTATGTAATT-----ATATATTATATATATTAATAATTATATTCGTATATATCATATTAT	33807
QY	1473	agtaataaagaaacgaggtagcccaacaaagtgataattgtgagggtgtgactcttg	1532
DB	33806	ATAATTATATAAATATATATGTTATAAATATATACATATATAGCTTATATGTATATATTAT	33747
QY	1533	tcggtgcataaaatgaaacccccaaactgttgatattdgtcgactgctccgctcacat	1592
DB	33746	ATAATTATATATATTAATAATTATATATAATAATAATAATATATATATAAATTTATAATAT	33687
QY	1593	tgaataaataagatgtctctttataacgcttgtctatgcggtattaccatatgctcact	1652
DB	33686	AAATATATATTTATATATAATATATAATTTATACATACATGCTGTTTAAATATAATAATATA	33627
QY	1653	agaatgggacaatgaaatttaatatatctgtcatgtgtgtgggtgattccaatttaattgt	1712
DB	33626	TAATTATATATAACTGTATATATAATAATACTATAATAT-----AATTATATAATAATTAT	33574
QY	1713	atcgtaaatggtaggacacatcatgctacacattatcatcatcactggccaactcgtg	1772
DB	33573	ATTATATATTTATATAATAATATATAAATTTATTTATTTATTTATAGTATATAATAATAATAAT	33514
QY	1773	tcaatgtgtttctctcccatgaattccaattgctgaagaaaaattaccaccttaaaatg	1832
DB	33513	TTATTATATTATATAGTATATATAATAATAATAATAATAATAATAATATGTAATATAATAAT	33454
QY	1833	ttatcccttgcaacatttcacatcaatttttaaacaattttaccattggaaaaacaca	1892
DB	33453	ATAATATATAATATATAATAATAATAATTTATATAAATAATTATCATATTTATTTATA	33394
QY	1893	tacatatccaatcaattttttgtcatttttcaaaaaactaaaccaaagaaacttagaat	1952

Db	33993	ATTATATATTATATATATAGTTATGCTAAATATATATATATATATAATAATATATAT	333934
Qy	1953	tttataattatagacaaattttcaaaaatactcagttctcaaccacftcaataattcaca	2012
Db	33333	TATATTATATAAATTAATATATATAAATTAATATAAATTAATACCATATATATATA	33274
Qy	2013	atttccaaaaccttgcaaaaacatcacaaacctcagaaaccttggatttaaatcctaataa	2072
Db	33273	ATTATATTATATAAATTAATAAATATATTACATATATAATTTATAAATAATATAAATA	33214
Qy	2073	aagcaataaatgatgatctcaacaatcacccatatagttgatgataaatcatgatgcag	2132
Db	33213	TAATTTATAAATACATATATATATATAAATTTATAAATAATA---TAAATTTATTTTAAATATAT	33157
Qy	2133	caatacaccttaatttggtaagcattataaagcgacacaactctattaacaccggtaattca	2192
Db	33156	ATAATTTATAAATATACAAATTTTATATATATAATATATATATGTAATATATACTATATAAA	33097
Qy	2193	acaacgggtgtgcgagttcatgtttcttccaaactcttttccctttcccttactttat	2252
Db	33096	TTATATTATAATTTTATATAATAAATATATAACATATAAATTTATATATAAATATATAAT	33037
Qy	2253	ttatttctctacttaccottttctactaataataatactatctctcttgaaacctcttttga	2312
Db	33036	AATTATATAACATATAAATTTATATAAATATAAATAAATTAATATAACATAAATAATTG	32977
Qy	2313	tctt	2316
Db	32976	TATT	32973

Search completed: April 3, 2002, 05:43:00
Job time: 5895 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: April 3, 2002, 04:04:45 ; Search time 1999.71 Seconds
(without alignments)
98.997 Million cell updates/sec

Title: US-09-700-187-1

Perfect score: 12

Sequence: 1 ggattttacagt 12

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_om.*

20: em_or.*

21: em_ov.*

22: em_pat.*

23: em_ph.*

24: em_pl.*

25: em_ro.*

26: em_sts.*

27: em_sy.*

28: em_un.*

29: em_vi.*

30: em_htgo_hum.*

31: em_htgo_inv.*

32: em_htgo_rod.*

33: em_htg_hum.*

34: em_htg_inv.*

35: em_htg_rod.*

36: em_htg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	12	100.0	22	6	AX004331	Sequence
C 2	12	100.0	145	5	CHRMIL08	K03266 Chicken c-m
C 3	12	100.0	164	11	G44227	G44227 WIAF-3873-S
C 4	12	100.0	164	11	G59874	G59874 SHGC-130652
C 5	12	100.0	215	9	HSA006490	AJ006490 Homo sapi
C 6	12	100.0	220	9	APU09224	U09224 Alouatta pa
C 7	12	100.0	222	11	G48131	G48131 SMSS31 Mous
C 8	12	100.0	228	14	AU038679	U38679 Avian infec
C 9	12	100.0	231	14	AU038677	U38677 Avian infec
C 10	12	100.0	231	14	AU038678	U38678 Avian infec
C 11	12	100.0	231	14	AU038681	U38681 Avian infec
C 12	12	100.0	285	11	AU025328	AU025328 Rattus no
C 13	12	100.0	366	14	HPCTH13P30	D28849 Hepatitis C
C 14	12	100.0	368	14	IBR243261	AJ243261 Infectiou
C 15	12	100.0	388	11	G30656	G30656 human STS S
C 16	12	100.0	454	11	G56069	G56069 SHGC-101332
C 17	12	100.0	463	11	HUMUT5361	L18645 Human chrom
C 18	12	100.0	475	11	G51167	G51167 SHGC-80775
C 19	12	100.0	679	6	A86412	A86412 Sequence 10
C 20	12	100.0	679	6	ARI55905	ARI55905 Sequence
C 21	12	100.0	679	6	E66430	E66430 Genome DNA
C 22	12	100.0	681	2	AC061540	AC061540 Giardia i
C 23	12	100.0	724	6	AX187114	AX187114 Sequence
C 24	12	100.0	725	5	AF231326	AF231326 Nyctea sc
C 25	12	100.0	725	5	AF231331	AF231331 Asio otus
C 26	12	100.0	740	2	AC038063	AC038063 Giardia i
C 27	12	100.0	776	2	AC015328	AC015328 Drosophil
C 28	12	100.0	802	8	AF363433	AF363433 Fusarium
C 29	12	100.0	803	8	AF363432	AF363432 Fusarium
C 30	12	100.0	919	9	AF069970	AF069970 Nasalis l
C 31	12	100.0	920	2	AC082280	AC082280 Giardia i
C 32	12	100.0	930	2	AF105062	AF105062 Phytolacc
C 33	12	100.0	935	9	BC008190	BC008190 Homo sapi
C 34	12	100.0	954	2	AC081006	AC081006 Giardia i
C 35	12	100.0	980	2	AC032864	AC032864 Giardia i
C 36	12	100.0	1000	2	AC062314	AC062314 Giardia i
C 37	12	100.0	1023	2	AC064712	AC064712 Giardia i
C 38	12	100.0	1080	2	AC062313	AC062313 Giardia i
C 39	12	100.0	1097	2	AC039640	AC039640 Giardia i
C 40	12	100.0	1138	6	AR083280	AR083280 Sequence
C 41	12	100.0	1138	6	AX099541	AX099541 Sequence
C 42	12	100.0	1188	6	AX011663	AX011663 Sequence
C 43	12	100.0	1307	3	AF216214	AF216214 Drosophil
C 44	12	100.0	1308	3	AF318611	AF318611 Caenorhab
C 45	12	100.0	1323	8	AF238237	AF238237 Arabidops

ALIGNMENTS

RESULT 1	AX004331/c	AX004331	22 bp	DNA	PAT	24-AUG-2000
LOCUS	Sequence 83 from Patent WO9919492.					
DEFINITION	AX004331					
ACCESSION	AX004331.1	GI:9927813				
VERSION						
KEYWORDS	synthetic construct.					
SOURCE	synthetic construct.					
ORGANISM	artificial sequence.					
REFERENCE	1 (bases 1 to 22)					
AUTHORS	Betzner, A.S. and Doutriaux, M.P.					
TITLE	Methods for obtaining plant varieties					
JOURNAL	Patent: WO 9919492-A 83 22-APR-1999.					
FEATURES	BETZNER ANDREAS STEFAN (AU); DOUTRIAUX MARIE PASCALE (FR)					
source	Location/Qualifiers					
	1..22					
	/organism="synthetic construct"					
	/db_xref="taxon:32630"					
	/note="Reverse primer for PCR amplification of NGA249 SSLP"					

```

BASE COUNT      7 a      7 c      3 g      5 t
ORIGIN
marker in Arabidopsis thaliana subspecies"

Query Match      100.0%; Score 12; DB 6; Length 22;
Best Local Similarity 100.0%; Pred. No. 5.9e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
   |||||
Db 21 GGATTTTACAGT 10

RESULT 2
CHRM108/c      145 bp      DNA      VRT      28-APR-1993
LOCUS      Chicken c-mil proto-oncogene, exon 8.
DEFINITION      K03266
ACCESSION      K03266.1 GI:212316
VERSION      c-myc proto-oncogene; mil oncogene; proto-oncogene.
KEYWORDS      8 of 11
SEGMENT      Chicken DNA, clones lambda c-mil-4 and lambda c-mil-7.
SOURCE      Gallus gallus
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
              Phasianinae; Gallus.
REFERENCE      1 (bases 1 to 145)
AUTHORS      Jansen,H.W. and Bister,K.
TITLE      Nucleotide sequence analysis of the chicken gene c-mil, the
              progenitor of the retroviral oncogene v-mil
JOURNAL      Virology 143, 359-367 (1985)
MEDLINE      86045899
COMMENT      See segment 1.
FEATURES
    source
        1..145
            Location/Qualifiers
                /organism="Gallus gallus"
                /db_xref="taxon:9031"
                <!--.20
                21..139
                    /note="c-mil intron H"
                    /note="c-mil protein, (putative); putative"
                    /number=8
                    /number=140..>145
                    /note="c-mil intron I"
                37 a      27 c      38 g      43 t
                About 900 kb after segment 7.

Query Match      100.0%; Score 12; DB 5; Length 145;
Best Local Similarity 100.0%; Pred. No. 5.2e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
   |||||
Db 85 GGATTTTACAGT 74

RESULT 3
G44227/c      164 bp      DNA      STS      28-JAN-1999
LOCUS      WIAF-3873-STS Human Thudson SANGER Homo sapiens STS genomic,
DEFINITION      sequence tagged site.
ACCESSION      G44227
VERSION      G44227.1 GI:4193144
KEYWORDS      STS.
SOURCE      human.
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 164)
AUTHORS      Wang,D.G., Fan,J.B., Siao,C.J., Berno,A., Young,P., Sapolsky,R.,
              Ghandour,G., Perkins,N., Winchester,E., Spencer,J., Kruglyak,L.,

Stein,L., Hsie,L., Topaloglou,T., Hubbell,E., Robinson,E.,
Mittmann,M., Morris,M.S., Shen,N., Kilburn,D., Rioux,J.,
Nusbaum,C., Rozen,S., Hudson,S., Lipshutz,R., Chee,M. and
Lander,E.S.
Large-scale identification, mapping, and genotyping of
single-nucleotide polymorphisms in the human genome
Science 280 (5366), 1077-1082 (1998)
98248615
Synonyms: stSG28550
Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
Whitehead Institute for Biomedical Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu
Primer A: TTGTATAGCTGGTGAGGTGC
Primer B: TGGTGCCTTATCCCAACATT
STS size: 164
PCR Profile:
    Presoak: 94 degrees C for 4.00 minutes
    Denaturation: 94 degrees C for 50.0 seconds
    Annealing: 58 degrees C for 1.50 minutes
    Polymerization: 72 degrees C for 1.00 minutes
    PCR Cycles: 30
    Thermal Cycler: custom built by IAS, Costar, Cambridge MA

Protocol:
    Template: 10 ng
    Primer: each 5 pM
    dNTPs: 4 nM
    Taq Polymerase: 0.5 U
    Total Vol: 20 uL

Buffer:
    Mg2+: 1.5 mM
    KCl: 50 mM
    Tris-HCl: 10 mM
    Gelatin: .001 %.
    Location/Qualifiers
        1..164
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /map="68.70 cR from top of Chr19 linkage group"
            /clone_lib="Human Thudson SANGER"
            /note="human STS created from EST in the Sanger database"
        STS
        primer_bind 1..164
        primer_bind 1..21
        primer_bind complement(145..164)
    BASE COUNT 43 a 25 c 36 g 59 t 1 others

Query Match      100.0%; Score 12; DB 11; Length 164;
Best Local Similarity 100.0%; Pred. No. 5.2e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
   |||||
Db 88 GGATTTTACAGT 77

RESULT 4
G59874/c      164 bp      DNA      STS      30-MAR-2000
LOCUS      SHGC-130652 Human Homo sapiens STS genomic, sequence tagged site.
DEFINITION      G59874
ACCESSION      G59874
VERSION      G59874.1 GI:6125193
KEYWORDS      STS.
SOURCE      human.
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 164)

```



```

Db 81 GGATTTTACAGT 70

RESULT 7
G48131/c
LOCUS      222 bp      DNA      STS      09-APR-1999
DEFINITION Mouse Eggen Mus musculus STS genomic, sequence tagged site.
ACCESSION  G48131
VERSION     G48131.1  GI:4494739
KEYWORDS   STS.
SOURCE     house mouse.
ORGANISM   Mus musculus.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 222)
AUTHORS   Desilva,U., Massa,H., Trask,B.J. and Green,E.D.
TITLE     Comparative Mapping of the Region of Human Chromosome 7 Deleted in
            William's Syndrome
JOURNAL   Unpublished (1999)
COMMENT   Synonyms: 92N10R
            Contact: Eric D. Green
            Genome Technology Branch
            National Human Genome Research Institute/NIH
            49 Convent Dr., MSC4431, Bldg. 49, Rm. 2A08, Bethesda, MD 20892
            Tel: 3014020201
            Fax: 3014024735
            Email: egreen@nhgri.nih.gov
            Primer A: TGCCTCTAGTCACAGGATGTGC
            Primer B: AGATGCTTGGTGTGGTAGGGTG
            STS size: 222
            PCR Profile:
                Presoak: 0 degrees C for 0.00 minute(s)
                Denaturation: 92 degrees C for 0.17 minute(s)
                Annealing: 58 degrees C for 1.00 minute(s)
                Polymerization: 72 degrees C for 1.00 minute(s)
                PCR Cycles: 35
            Thermal Cycler: PerkinElmer 9600
            Protocol:
                Template: 30-100 ng
                Primer: each 1 uM
                dNTPs: each 200 uM
                Taq Polymerase: 0.05 units/ul
                Total Vol: 10 ul

Buffer:
MgCl2: 1.5 mM
KCl: 100 mM
Tris-HCl: 10 mM
NH4Cl: 5 mM
pH: 8.6

The sequence for this STS was derived from a single sequencing
read.

FEATURES             source
    STS               1..222
                        /organism="Mus musculus"
                        /db_xref="taxon:10090"
                        /clone_lib="Mouse EGgen"
    primer_bind       1..22
    primer_bind       1..222
    primer_bind       complement(201..222)
BASE COUNT          64 a 61 c 42 g 53 t 2 others
ORIGIN
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2370
2371
2372
2373
2374
2375
2376
2377
2378
2379
2380
2381
2382
2383
2384
2385
2386
2387
2388
2389
2390
2391
2392
2393
2394
2395
2396
2397
2398
2399
2400
2401
2402
2403
2404
2405
2406
2407
2408
2409
2410
2411
2412
2413
2414
2415
2416
2417
2418
2419
2420
2421
2422
2423
2424
2425
2426
2427
2428
2429
2430
2431
2432
2433
2434
2435
2436
2437
2438
2439
2440
2441
2442
2443
2444
2445
2446
2447
2448
2449
2450
2451
2452
2453
2454
2455
2456
2457
2458
2459
2460
2461
2462
2463
2464
2465
2466
2467
2468
2469
2470
2471
2472
2473
2474
2475
2476
2477
2478
2479
2480
2481
2482
2483
2484
2485
2486
2487
2488
2489
2490
2491
2492
2493
2494
2495
2496
2497
2498
2499
2500
2501
2502
2503
2504
2505
2506
2507
2508
2509
2510
2511
2512
2513
2514
2515

```

TITLE Direct Submission
JOURNAL Submitted (17-OCT-1995) Ching-Ho Wang, National Taiwan University,
Vetinary Medicine, 142 Chou-san Rd., Taipei, 106, Taiwan
FEATURES Location/Qualifiers
source 1..231
/organism="Avian infectious bronchitis virus"
/strain="A1211"
/db_xref="taxon:11120"
gene 1..231
/gene="SI"
CDS <1..>231
/gene="SI"
/note="hypervariable region 1; N-terminus"
/codon_start=2
/product="spike protein"
/protein_id="AAB47435.1"
/db_xref="GI:1055302"
/translation="GWLHGGAYAVVVSSTNNAGSACTVGTIRGDRVNVNASSIA
MTAPVQGMQWSKQFCTAHCNFSDFTVFVTH"
BASE COUNT 56 a 40 c 55 g 80 t
ORIGIN

Query Match 100.0%; Score 12; DB 14; Length 231;
Best Local Similarity 100.0%; Pred. No. 5.1e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
|||||
Db 205 GGATTTTACAGT 216

RESULT 10
AU38678
LOCUS 231 bp RNA VRL 14-FEB-1997
DEFINITION Avian infectious bronchitis virus A1960 spike protein (SI) gene,
hypervariable region 1, partial cds.
ACCESSION U38678
VERSION 1
KEYWORDS
SOURCE
ORGANISM
Avian infectious bronchitis virus.
Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
Coronaviridae; Coronavirus.
REFERENCE 1 (sites)
AUTHORS Wang, C.H. and Tsai, C.T.
TITLE Genetic grouping for the isolates of avian infectious bronchitis
virus in Taiwan
JOURNAL Arch. Virol. 141 (9), 1677-1688 (1996)
MEDLINE 97049060
REFERENCE 2 (bases 1 to 231)
AUTHORS Wang, C.
TITLE Direct Submission
JOURNAL Submitted (17-OCT-1995) Ching-Ho Wang, National Taiwan University,
Vetinary Medicine, 142 Chou-san Rd., Taipei, 106, Taiwan
FEATURES Location/Qualifiers
source 1..231
/organism="Avian infectious bronchitis virus"
/strain="A1955"
/db_xref="taxon:11120"
gene 1..231
/gene="SI"
CDS <1..>231
/gene="SI"
/note="hypervariable region 1; N-terminus"
/codon_start=2
/product="spike protein"
/protein_id="AAB47439.1"
/db_xref="GI:1055310"
/translation="GWLHGGAYAVVNVSSQTNNAGTAGCTVGIISGDRVVNASSIA
MSAPVQGMQWSKQFCTAHCNFSDFTVFVTH"
BASE COUNT 57 a 40 c 55 g 79 t
ORIGIN

Query Match 100.0%; Score 12; DB 14; Length 231;
Best Local Similarity 100.0%; Pred. No. 5.1e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
|||||
Db 205 GGATTTTACAGT 216

RESULT 12
AU025328
LOCUS 285 bp DNA STS 02-MAR-1999
DEFINITION Rattus norvegicus, OTSUKA clone, OT57.23/887g03, microsatellite
sequence, sequence tagged site.
ACCESSION AU025328
VERSION AU025328.1 GI:4515251
KEYWORDS STS.
SOURCE Rattus norvegicus DNA, clone:OT57.23/887g03.
ORGANISM Rattus norvegicus

TITLE Direct Submission
JOURNAL Submitted (17-OCT-1995) Ching-Ho Wang, National Taiwan University,
Vetinary Medicine, 142 Chou-san Rd., Taipei, 106, Taiwan
FEATURES Location/Qualifiers
source 1..231
/organism="Avian infectious bronchitis virus"
/strain="A1211"
/db_xref="taxon:11120"
gene 1..231
/gene="SI"
CDS <1..>231
/gene="SI"
/note="hypervariable region 1; N-terminus"
/codon_start=2
/product="spike protein"
/protein_id="AAB47435.1"
/db_xref="GI:1055302"
/translation="GWLHGGAYAVVVSSTNNAGSACTVGTIRGDRVNVNASSIA
MTAPVQGMQWSKQFCTAHCNFSDFTVFVTH"
BASE COUNT 56 a 40 c 55 g 80 t
ORIGIN

Query Match 100.0%; Score 12; DB 14; Length 231;
Best Local Similarity 100.0%; Pred. No. 5.1e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
|||||
Db 205 GGATTTTACAGT 216

RESULT 11
AU38681
LOCUS 231 bp RNA VRL 14-FEB-1997
DEFINITION Avian infectious bronchitis virus A1955 spike protein (SI) gene,
hypervariable region 1, partial cds.
ACCESSION U38681
VERSION 1
KEYWORDS
SOURCE
ORGANISM
Avian infectious bronchitis virus.
Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
Coronaviridae; Coronavirus.
REFERENCE 1 (sites)
AUTHORS Wang, C.H. and Tsai, C.T.
TITLE Genetic grouping for the isolates of avian infectious bronchitis
virus in Taiwan
JOURNAL Arch. Virol. 141 (9), 1677-1688 (1996)
MEDLINE 97049060
REFERENCE 2 (bases 1 to 231)
AUTHORS Wang, C.
TITLE Direct Submission
JOURNAL Submitted (17-OCT-1995) Ching-Ho Wang, National Taiwan University,
Vetinary Medicine, 142 Chou-san Rd., Taipei, 106, Taiwan
FEATURES Location/Qualifiers
source 1..231
/organism="Avian infectious bronchitis virus"
/strain="A1955"
/db_xref="taxon:11120"
gene 1..231
/gene="SI"
CDS <1..>231
/gene="SI"
/note="hypervariable region 1; N-terminus"
/codon_start=2
/product="spike protein"
/protein_id="AAB47439.1"
/db_xref="GI:1055310"
/translation="GWLHGGAYAVVNVSSQTNNAGTAGCTVGIISGDRVVNASSIA
MSAPVQGMQWSKQFCTAHCNFSDFTVFVTH"
BASE COUNT 57 a 40 c 55 g 79 t
ORIGIN

Query Match 100.0%; Score 12; DB 14; Length 231;
Best Local Similarity 100.0%; Pred. No. 5.1e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
|||||
Db 205 GGATTTTACAGT 216

RESULT 12
AU025328
LOCUS 285 bp DNA STS 02-MAR-1999
DEFINITION Rattus norvegicus, OTSUKA clone, OT57.23/887g03, microsatellite
sequence, sequence tagged site.
ACCESSION AU025328
VERSION AU025328.1 GI:4515251
KEYWORDS STS.
SOURCE Rattus norvegicus DNA, clone:OT57.23/887g03.
ORGANISM Rattus norvegicus

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
AUTHORS
Watanabe,T.K., Hishigaki,H., Kanemoto,N., Miyakita-Mizoguchi,A.,
Oga,K., Okuno,S., Ono,T., Tsuji,A., Hayashi,H., Adachi,M.,
Yamasaki,Y., Iriye,Y., Takahashi,E., Takagi,Y., Nakamura,Y. and
Tanigami,A.
TITLE
The large-scale mapping of rat microsatellite markers
JOURNAL
Unpublished (1998)
REFERENCE
2 (bases 1 to 285)
AUTHORS
Watanabe,T.K.
TITLE
Direct Submission
JOURNAL
Submitted (24-JUL-1998) to the DDBJ/EMBL/GenBank databases. Takeshi
K Watanabe, Otsuka Pharmaceutical Co., Ltd., Otsuka GEN Research
Institute, 463-10, Kagasuno, Kawachi-cho, Tokushima, Tokushima
771-0192, Japan (E-mail:watanabe@otsuka.genome.ad.jp,
Tel:81-886-65-2888, Fax:81-886-37-1035)
FEATURES
source
1. .285
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="OF57.23/887g03"
/notes="OF57.23/887g03P-5'-CCTTATTCCATGGGTACCTGA-3',
OF57.23/887g03R-5'-CGATCAGGTACTGCCATACC-3'"
BASE COUNT
114 a 80 c 35 g 56 t
ORIGIN
1 ggattttacagt 12
|||||
Db 205 GGATTTTACAGT 216

Query Match 100.0%; Score 12; DB 11; Length 285;
Best Local Similarity 100.0%; Pred. No. 5e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
|||||
Db 205 GGATTTTACAGT 216

RESULT 13
HPCTH13P30 366 bp RNA VRL 07-FEB-1999
LOCUS
Hepatitis C virus RNA for NS5, partial sequence.
DEFINITION
D28849
ACCESSION
D28849.1 GI:1009306
VERSION
NS5; TH13PI.
KEYWORDS
Hepatitis C virus (strain:TH13) RNA.
SOURCE
Hepatitis C virus
ORGANISM
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
REFERENCE
1 (bases 1 to 366)
AUTHORS
Sugiyama,K.
TITLE
Direct Submission
JOURNAL
Submitted (05-MAR-1994) to the DDBJ/EMBL/GenBank databases. Kazuo
Sugiyama, National Cancer Center Research Institute, Virology
Division, 5-1-1, Tsukiji, Chuo-ku, Tokyo 104, Japan
(Tel:03-3542-2511(ex.4701), Fax:03-3543-2181)
REFERENCE
2 (bases 1 to 366)
AUTHORS
Sugiyama,K.
JOURNAL
Unpublished (1995)
REFERENCE
3 (sites)
AUTHORS
Sugiyama,K., Kato,N., Nakazawa,T., Yonemura,Y., Phornphutkul,K.,
Kunakorn,M., Petchclai,B. and Shimotohno,K.
TITLE
Novel genotypes of hepatitis C virus in Thailand
JOURNAL
J. Gen. Virol. 76 (Pt 9), 2323-2327 (1995)
MEDLINE
96005056
FEATURES
Location/Qualifiers
source
1. .366
/organism="Hepatitis C virus"
/strain="TH13"
/db_xref="taxon:111103"
1. .366
/gene="TH13PI"
<1. .>366

gene
CDS

```

```

/gene="TH13PI"
/codon_start=1
/product="NS5"
/protein_id="BAA05996.1"
/db_xref="GI:1669629"
/translation="FPVNSWLGNIIMYAPTIVRWVLMTHFFGILQSQETLHKALDFD
MTGVYSITPLDLPLQIIRLHGMAAFSLHGYSFGELNIRVAASLRKLGAPPLRAWRHRA
RAVRAKLIAQGGKAATCGRY"
BASE COUNT
89 a 100 c 94 g 83 t
ORIGIN
1 ggattttacagt 12
|||||
Db 86 GGATTTTACAGT 97

Query Match 100.0%; Score 12; DB 14; Length 366;
Best Local Similarity 100.0%; Pred. No. 4.9e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
|||||
Db 86 GGATTTTACAGT 97

RESULT 14
IBR243261 368 bp RNA VRL 31-JAN-2000
LOCUS
Infectious bronchitis virus S gene, strain 624/I (368bp).
DEFINITION
AJ243261
ACCESSION
AJ243261.1 GI:5262733
VERSION
S gene.
KEYWORDS
Avian infectious bronchitis virus.
SOURCE
Avian infectious bronchitis virus
ORGANISM
Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
Coronaviridae; Coronavirus.
REFERENCE
1 (bases 1 to 368)
AUTHORS
Capua,I., Minta,Z., Karpinska,E., Mawditt,K., Britton,P.,
Cavanagh,D. and Gough,R.E.
TITLE
Co-circulation of four types of infectious bronchitis virus (793/B,
624/I, B1648 and Massachusetts)
JOURNAL
Avian Pathol. 28, 587-592 (1999)
REFERENCE
2 (bases 1 to 368)
AUTHORS
Capua,I.
TITLE
Direct Submission
JOURNAL
Submitted (18-JUN-1999) Capua I., Istituto Zooprofilattico delle
Venezie, Via Romea 14/A, 35020, Legnaro, Padova, Italy
FEATURES
Location/Qualifiers
source
1. .368
/organism="Avian infectious bronchitis virus"
/strain="624/I"
/db_xref="taxon:11120"
1. .368
/gene="S"
<1. .>368
/gene="S"
/codon_start=1
/protein_id="CAB45685.1"
/db_xref="GI:5262734"
/db_xref="SPTREMBL:Q9WML2"
/translation="NAGTAQOCTAGAIKSNKFSASSVAMTAPGSGMQMSTNOFCTAH
CNFTDFTVFVTHCFKSGAGQCPLTGTFLPSGVIRVSAMRKGNFSLFYNLTVSVTKYPTF
KSLQCVNHNHVSFLNGDLVFT"
BASE COUNT
92 a 73 c 75 g 128 t
ORIGIN
1 ggattttacagt 12
|||||
Db 144 GGATTTTACAGT 155

Query Match 100.0%; Score 12; DB 14; Length 368;
Best Local Similarity 100.0%; Pred. No. 4.9e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
|||||
Db 144 GGATTTTACAGT 155

RESULT 15

```

G30656
LOCUS 388 bp DNA STS 05-OCT-1996
DEFINITION human STS SHGC-37261, sequence tagged site.
ACCESSION G30656
VERSION G30656.1 GI:1594207
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 388)
AUTHORS Myers,R.M.
JOURNAL Unpublished (1996)
COMMENT
Contact: Richard M. Myers
Stanford Human Genome Center (SHGC)
Stanford University School of Medicine
Department of Genetics, M-344, Stanford, CA 94305, USA
Tel: 4157259687
Fax: 4157259689
Email: myers@shgc.stanford.edu
Primer A: TAAAGCAACCACTAATGTCCTGG
Primer B: CTTGCAGTTATTGCTAGATGG
STS size: 128
PCR Profile:
Initial incubation: 94 degrees C for 90 seconds
Denaturation: 94 degrees C for 15 seconds
Annealing: 62 degrees C for 23 seconds
Polymerization: 72 degrees C for 30 seconds
PCR Cycles: 30
Thermal Cycler: Perkin Elmer 9600
Protocol:
Template: 25 ng
Primer: each 1 uM
dNTPs: each 200 uM
Taq Polymerase: 0.05 units/ul
Total Vol: 10 ul
Buffer:
MgCl2: 2.5 mM
KCl: 50 mM
Tris-HCl: 20 mM
pH: 8.3
Prepared with primer pairs provided by Sandoz, derived from H66941
-- Washington University/Merck EST sequence.
FEATURES
source 1..388
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="13"
STS 10..137
primer_bind 10..33
primer_bind complement(115..137)
BASE COUNT 130 a 74 c 68 g 116 t
ORIGIN
Query Match 100.0%; Score 12; DB 11; Length 388;
Best Local Similarity 100.0%; Pred. No. 4.9e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ggattttacagt 12
|||||
Db 152 GGATTTTACAGT 163

Search completed: April 3, 2002, 05:13:00
Job time: 4095 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2002, 04:39:12 ; Search time 2120.21 seconds
(without alignments)
471.348 Million cell updates/sec

Title: US-09-700-187-2

Perfect score: 93

Sequence: 1 aaaagtaacacatatattga.....ttacagtaataaagaacga 93

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

- 1: em_estfun:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estom:*
- 5: em_estpi:*
- 6: em_estba:*
- 7: em_estro:*
- 8: em_estov:*
- 9: em_htc:*
- 10: gb_estl:*
- 11: gb_est2:*
- 12: gb_htc:*
- 13: gb_gss:*
- 14: em_gss_fun:*
- 15: em_gss_hum:*
- 16: em_gss_inv:*
- 17: em_gss_pln:*
- 18: em_gss_pro:*
- 19: em_gss_rod:*
- 20: em_gss_vrt:*
- 21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
c 1	37.8	40.6	357	10	AA169463
c 2	36.4	39.1	320	13	BH036705
c 3	36.4	39.1	555	13	BH102413
c 4	35.4	38.1	489	13	CNS03DHC
c 5	35.4	38.1	1714	11	BG024018
c 6	34.8	37.4	521	11	C90222
c 7	34.6	37.2	631	10	AV405634
c 8	34.4	37.0	442	13	AQ776172
c 9	34	36.6	846	13	AQ743241
c 10	33.6	36.1	515	13	AQ764101
c 11	33.4	35.9	587	10	AJ388941
c 12	33.2	35.7	545	13	AQ545066

c 13	33	35.5	545	10	AV739051
c 14	33	35.5	547	13	AZ648187
c 15	33	35.5	549	13	TA195A02P
c 16	33	35.5	558	13	TA12F02P
c 17	33	35.5	947	11	BG025969
c 18	33	35.5	1082	13	CNS05FLS
c 19	32.8	35.3	321	13	B56628
c 20	32.6	35.1	924	13	B12729
c 21	32.6	35.1	1190	13	B10407
c 22	32.4	34.8	223	10	BE030084
c 23	32.4	34.8	631	13	AZ334849
c 24	32.4	34.8	669	13	AZ516773
c 25	32.4	34.8	1062	13	CNS00LU6
c 26	32.2	34.6	964	13	CNS005OV
c 27	32	34.4	534	11	BG555297
c 28	32	34.4	907	11	BE873019
c 29	31.8	34.2	323	10	AV017830
c 30	31.8	34.2	596	10	BE217602
c 31	31.8	34.2	945	13	CNS0605W
c 32	31.8	34.2	1099	10	AL536986
c 33	31.8	34.2	1101	13	CNS016VF
c 34	31.6	34.0	216	10	AL699725
c 35	31.6	34.0	237	11	F02028
c 36	31.6	34.0	252	11	N73114
c 37	31.6	34.0	343	10	AI038703
c 38	31.6	34.0	346	11	R92539
c 39	31.6	34.0	350	10	AW512010
c 40	31.6	34.0	356	10	AI200734
c 41	31.6	34.0	370	10	AW504912
c 42	31.6	34.0	393	10	AW010162
c 43	31.6	34.0	414	11	H47009
c 44	31.6	34.0	417	11	R52980
c 45	31.6	34.0	420	11	N55097

ALIGNMENTS

RESULT 1
AA169463/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AA169463 357 bp mRNA EST 09-NOV-1997
z084h08.r1 Stratagene ovarian cancer (#937219) Homo sapiens CDNA
clone IMAGE:593631 5', mRNA sequence.

AA169463.1 GI:1748411

human.

EST.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 357)

Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,

Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin

, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,

White, Y., Wylie, T., Waterston, R. and Wilson, R.

WashU-NCI human EST Project

Unpublished (1997)

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 1082 Std Error: 0.00

Seq primer: -28M13 rev2 from Amersham

High quality sequence stop: 251.

Location/Qualifiers

1..357

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:593631"

```

/clone_lib="Stratagene ovarian cancer (#937219)"
/sex="female"
/dev_stage="adult, 64 years"
/lab_host="SOLR (Kanamycin resistant)"
/notes="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
; Cloned unidirectionally. Primer: Oligo dT. Papillary
serous carcinoma, isolated from ascites, 64 year old
caucasian. Average insert size: 0.8 kb; Uni-ZAP XR Vector;
-5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3' adaptor
sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3'"
BASE COUNT      104 a      48 c      57 g      146 t
ORIGIN
Query Match      40.6%; Score 37.8; DB 10; Length 357;
Best Local Similarity 64.0%; Pred. No. 37;
Matches 57; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1 aaagtaacacatatatttgataaaatttatactaaactatttcttagtactgttaac 60
||||| ||||| || ||||||| ||||||| ||||| ||||| || ||
Db 216 AAACAACACAGCTAAGCAAAATTAATTAATAAATAAACAATTTCTAAGTAGATGAATATA 157
QY 61 atgtctgaggtttacagtataaagaa 89
||| ||||| ||||||| ||||||| ||||||| ||
Db 156 ATCACGATGTTTATACAGCTATAAGCA 128

RESULT 2
BH036705
LOCUS      BH036705      320 bp      DNA      GSS      17-JUL-2001
DEFINITION      BH036705      Mus musculus genomic clone RPCI-24-22406,
DNA sequence.
ACCESSION      BH036705
VERSION      BH036705.1 GI:14811846
KEYWORDS
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 320)
Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Akinret,B., Levins,M.,
Russell,D., de Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-24
Unpublished (1999)
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (http://www.tigr.org/bacpac/orderingframe.htm). BAC end
page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 224 row: 0 column: 6
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1..320
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-22406"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J"

FEATURES
source
Location/Qualifiers
1..555
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-229N23"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J"

```

```

BASE COUNT      102 a      58 c      47 g      113 t
ORIGIN
Query Match      39.1%; Score 36.4; DB 13; Length 320;
Best Local Similarity 64.0%; Pred. No. 77;
Matches 55; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 1 aaagtaacacatatatttgataaaatttatactaaactatttcttagtactgttaac 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 159 AAAGAAATACAAATATGACTCCCTTTTCTTAAACAATATTTTAGTAATCTTCTTC 218
QY 61 atgtctgaggtttacagtataa 86
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 219 TTCCTTAAGTTATTAGCAGCAATAGA 244

RESULT 3
BH102413
LOCUS      BH102413      555 bp      DNA      GSS      19-JUL-2001
DEFINITION      BH102413      Mus musculus genomic clone RPCI-24-229N23
DNA sequence.
ACCESSION      BH102413
VERSION      BH102413.1 GI:14928295
KEYWORDS      GSS.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 555)
Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-24
Unpublished (1999)
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (http://www.tigr.org/bacpac/orderingframe.htm). BAC end
page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 229 row: N column: 23
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1..555
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-229N23"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J"
DNA."
BASE COUNT      143 a      113 c      86 g      213 t
ORIGIN
Query Match      39.1%; Score 36.4; DB 13; Length 555;
Best Local Similarity 64.0%; Pred. No. 66;
Matches 55; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

```



```

Qy 1 aaaagtaacacatatatttgataaaatttactaaactattttctagtagtctgttaac 60
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 156 AAAGAAATACAAATATGACTCCCTTTTACTTAAACAATATTTTACTAATCTTCTC 215
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 61 atgtctgaggattttacagtaataaa 86
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 216 TTCTTAAAGTTATTAGCAGCAATAGA 241
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 4
CNS03DHC/c
LOCUS
DEFINITION CNS03DHC 489 bp DNA GSS 15-MAY-2000
Tetraodon nigroviridis genome survey sequence PUC-ori end of clone
017A03 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL239097.1 GI:7898232
VERSION AL239097.1
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
1 (bases 1 to 489)
Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizes,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Unpublished
2 (bases 1 to 489)
Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizes,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished
3 (bases 1 to 489)
Genoscope.
Direct Submission
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/tetraodon.
FEATURES
source
Location/Qualifiers
1..489
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="017A03"
/clone_lib="G"
/note="Genoscope sequence ID : CBG017AA02SP1-end :
PUC-ori"
BASE COUNT 243 a 21 c 105 g 85 t 35 others
ORIGIN

Query Match 38.1%; Score 35.4; DB 13; Length 489;
Best Local Similarity 64.6%; Pred. No. 1.1e+02;
Matches 51; Conservative 1; Mismatches 27; Indels 0; Gaps 0;

Qy 6 taacacattttgataaaatttactaaactattttctagtagtctgttaacatgc 65
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 97 TTAACATTTTTAAACAATCTATACAWAAATCTTTTCTCTTCTTCTTCTTCTTCTC 38
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 66 tgaggattttacagtaata 84
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 37 ATCAAAATTTAAATAATA 19
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 5

```

```

BG024018
LOCUS
DEFINITION BG024018 1714 bp mRNA EST 24-JAN-2001
602303237F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4394600 5',
mRNA sequence.
ACCESSION BG024018
VERSION BG024018.1 GI:12409159
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1714)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10090 row: j column: 09
High quality sequence stop: 30.
FEATURES
source
Location/Qualifiers
1..1714
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4394600"
/clone_lib="NIH_MGC_88"
/tissue_type="duodenal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: small intestine; Vector: pCMV-SPORT6;
Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
oligo-dT primed. Average insert size 1.767 Kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH_MGC Library."
BASE COUNT 919 a 209 c 457 g 129 t
ORIGIN

Query Match 38.1%; Score 35.4; DB 11; Length 1714;
Best Local Similarity 61.3%; Pred. No. 80;
Matches 57; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Qy 1 aaagtaacacatatatttgataaaatttactaaactattttctagtagtctgttaac 60
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 66 AAATAATAATATATATTATTAATATATATAATATAATATAATATATATATATTTT 125
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 61 atgtctgaggattttacagtaataaaagaaacga 93
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 126 TTTTATTTTATTTATTTATTTATTAATAATGAAGA 158
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 6
C90222
LOCUS
DEFINITION C90222 Dictyostelium discoideum SS (H.Urushi-hara) Dictyostelium
discoideum cDNA clone SSI395, mRNA sequence.
ACCESSION C90222
VERSION C90222.1 GI:3059842
KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
1 (bases 1 to 521)
Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
Yoshino,R., Mitra,B.N., Pl,M., Sato,T., Takemoto,K., Yasukawa,H.,
Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
Developmental cDNA in Dictyostelium discoideum
Unpublished (1998)
TITLE
JOURNAL

```

7

Qy 68 aggattttacagtaataaagaac 91
 ||| ||| ||| ||| ||| ||| |||
 Db 348 AGTAATATGCAGCCATAAACACAC 325

RESULT	9
AQ743241/c	
LOCUS	846 bp DNA GSS 16-JUL-1999
DEFINITION	HS_5388_B1_D10_SP6 RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=964 Col=19 Row=H, DNA sequence.
ACCESSION	AQ743241
VERSION	AQ743241.1 GI:5520763
KEYWORDS	GSS.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 846)
AUTHORS	Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.

JOURNAL
MEDLINE
COMMENT

Sequencing the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering/bac.htm>)
or from Research Genetics (<http://inforesgen.com>). BAC end Web Server:
<http://www.htsc.washington.edu>
Plate: 964 row: H column: 19
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 846.

RESULT 10
AQ764101

LOCUS	AQ764101	515 bp	DNA	GSS	28-JUL-1999
DEFINITION	HS_2001_AL_C12_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate-2001 Col=23 Row=M, DNA sequence.				
ACCESSION	AQ764101				
VERSION	AQ764101.1				
KEYWORDS	GI:5642217				
SOURCE	GSS.				
ORGANISM	human.				
	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.				
REFERENCE	1 (bases 1 to 515)				
AUTHORS	Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and Hood,L.				
TITLE	Sequencing-tagged connectors: A sequence approach to mapping and scanning the human genome				
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)				
MEDLINE	99380589				
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington				

```

tel: (206) 616-3887
fax: (206) 616-3887
Email: jwallaceu.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 2001 row: M column: 23
Seq primer: M13 Reverse
Class: BAC ends
High quality sequence stop: 515.
      location/Qualifiers
1. 515
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /clone="Plate=2001 Col=23 Row=M"
   /clone_lib="CIT Approved Human Genomic Sperm Library D"
   /sex="male"
   /note="Origin: sperm; Vector: pBelobAC11; BAC Clones in
E-Coll DH10B"
178 a      84 c      86 g      156 t      11 others
BASE COUNT
ORIGIN

```

Query Match	36.1%	Score 33.6;	DB 13;	Length 515;
Best Local Similarity	61.4%	Pred. No. 2.7e+02;		
Matches 54;	Conservative 0;	Mismatches 34;	Indels 0;	Gaps 0;

Qy 2	aagtaaacacacatatatttgatataaattattactcaaaaactattttctagctactgttaatca	61
Db 49	ACATTTCACATACATACATAGAATTTGTGTCAACACATTTGCTAGTACTAATTAATA	108
Qy 62	tgtctgaggattttcacagtaataaagaa	89
Db 109	ATTGTGTGATTGATATAGTAGACAGTA	136

RESULT 11				
AJ388941/c				
LOCUS	AJ388941	587 bp	mRNA	21-MAR-2000
DEFINITION	Medicago truncatula R108 Medicago truncatula cDNA clone			
	MTN0312, mRNA sequence.			

EST. 00541.1. 01.0004208
barrel medic.
Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids 1; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
1 (bases 1 to 587)
Gyorgyey,J., Vaubert,D., Jimenez-Zurdo,J.I., Charon,C., Troussard

TITLE L., Kondorosi A. and Kondorosi E.
JOURNAL Analysis of Medicago truncatula nodule expressed sequence tags
MEDLINE Mol. Plant Microbe Interact. 13 (1), 62-71 (2000)
COMMENT 20120379
Contact: Gyorgyey J
Institut des Sciences Vegetales
Centre National de la Recherche Scientifique (CNRS)
Av. de la Terrasse Bat.23, Gif-sur-Yvette, 91198, FRANCE
Additional information about these EST clones can be downloaded
from
<http://www.cnrs-gif.fr/isv/AK/index.html>. (URL provided by the
author).

FEATURES
source

FEATURES	source	Location/Qualifiers
		i. .587
		/organism="Medicago truncatula"
		/strain="R108"
		/db_xref="taxon:3880"
		/clone="MtNo312"
		/clone_lib="Medicago truncatula R108"
		/tissue_type="symbiotic root nodule"
		/dev_stage="developing young nodule"
		/notes="Vector: NOL1y lambda HybriZAP"
BASE COUNT	201 a	75 c 99 g 206 t
ORIGIN		6 others

Query Match	35.9%	Score 33.4;	DB 10;	Length 587;
Best Local Similarity	64.5%;	Pred. No. 2.9e+02;		
Matches 49; Conservative	0;	Mismatches 27;	Indels 0;	Gaps 0;

Qy 3 aagtaacacatatattgataaaatttattactaaaactattttcttagtactgtgaatcat 62
||| ||| ||| | ||||||| ||| ||||| ||||| ||| |||||||
Db 447 AAATTNCAAAATATTTTCATACTTATTATTAATAATATTTCTTAGTTCCTGGGAATCAT 388

Qy 63 gtctgaggatttaca 78
| | | | | | | | |
Db 387 ACATAAGGACTTTATA 372

RESULT	12
AQ545066/c	
LOCUS	AQ545066 545 bp DNA
DEFINITION	CITBI-EI-263409.TF CITBI-E1 Homo sapiens genomic clone 263409, DNA sequence.
ACCESSION	AQ545066
VERSION	AQ545066.1
KEYWORDS	GSS.
SOURCE	human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 545)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
Venet,J.C.
TITLE Use of BAC End Sequences from CalTech Libraries for Sequence-Ready
Map Building
JOURNAL Unpublished (1997)
COMMENT Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbswtigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21

FEATURES
SOURCE

```

class: One class.
FEATURES
  Location/Qualifiers
    1..545
      /organism="Homo sapiens"
      /db xref="taxon:9606"
  source

```

```

/clone="263409"
/clone_lib="CITBI-EI"
/sex="male"
/cell_type="sperm"
/notes="Vector: pBelobAC11; Site_1: EcoRI;
          Caltech Human BAC Library D"
185 a 96 c 87 g 177 t
BASE COUNT
ORIGIN

```

Query Match 35.7%; Score 33.2; DB 13; Length 545;
Best Local Similarity 61.6%; Pred. NO. 3.3e+02;
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

Qy	2	aaagtaacacataatttttgataaattattactaaaactattttctctagttacttctgtaataca	61
Db	184	ATAGTGAATTATTTTTCCTGAAAAATTTTGTCAAAAAATAGTCAATAATAATTGACAATCG	125
Qy	62	tgctgaggattttacagtaataaag	87
Db	124	AGTCTGAATAAAATACCATATAAAG	99

RESULT	13
AV739051/c	
LOCUS	AV739051 545 bp mRNA EST
DEFINITION	AV739051 CB Homo sapiens CDNA clone CBLAPG08 5', mRNA sequence.
ACCESSION	AV739051
VERSION	AV739051.1 GI:10856632
KEYWORDS	EST.
SOURCE	human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Butheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 545)
AUTHORS Zhang, Q., Ye, M., Wu, X., Gu, J., Huang, Q., Zhou, J., Shen, Y., Han, Z.,
 Chen, S., Mao, M. and Chen, Z.
TITLE Homo sapiens CB library cDNA clones
JOURNAL Homo sapiens CB library cDNA clones
COMMENT Unpublished (2000)
 Contact: Zhu Chen

Shanghai Institute of Hematology, Rui-Jin Hospital
197 Rui-Jin II Road, Shanghai 200025, P. R. China
Tel: 86-21-64740490
Fax: 86-21-64743206
Email: mbshi@ms.stn.sh.cn
This clone is available at Shanghai Hematology Institute in Shanghai.
Chinese National Human Genome Center at Shanghai
351 Guo Shou'ing Road, Zhangjiang Hi-Tech Park, Pudong.

```

FEATURES
source
1. .545
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CBLAPC08"
/clone_lib="CB"
/tissue_type="cord blood"
/cell_type="CD34+ hematopoietic stem/progenitor cell"
/lab_host="BM25.8"
/note="vector: pBluescript; Site_1: EcoRI; The insert is
cloned randomly with the EcoRI digestion"
171 a 118 c 124 g 130 t 2 others
BASE COUNT
ORIGIN

```

```
Query Match      35.5%; Score 33; DB 10; Length 545;
Best Local Similarity 60.7%; Pred. No. 3.6e+02;
Matches 54; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
```

QY 2 aaagtaacacatatatttgataaaatttatctaaaaactatcttagtactgtttaatca 61
||| | | ||| | |||| | | ||| | ||| | ||| |
Db 210 AAATTCTGATAAATTCGTAAATAAGTCTCAGGAGAACCAACAATAGACTTAGTCAATCA 151

QY 62 tgctgaqdaattttcacagtaataaaaaga 90

```

150 TGTGGAAACTTAAGAGTAATGAAGTAA 122
|||||
Db 150 TGTGGAAACTTAAGAGTAATGAAGTAA 122

RESULT 14
AZ648187 547 bp DNA GSS 14-DEC-2000
LOCUS IM0517001F Mouse 10kb plasmid UUGCLM library Mus musculus genomic
DEFINITION clone UUGCLM0517001 F, DNA sequence.
ACCESSION AZ648187
VERSION AZ648187.1 GI:11780403
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 547)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mamoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.
and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0517 row: 0 column: 01
Seq primer: CGTTGTAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 547.
FEATURES
    source
        1..547
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="UUGCLM0517001"
            /clone_lib="Mouse 10kb plasmid UUGCLM library"
            /sex="Male"
            /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
            /note="vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gi14732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT 209 a 86 c 67 g 185 t
ORIGIN

Query Match 35.5%; Score 33; DB 13; Length 547;
Best Local Similarity 60.7%; Pred. No. 3.6e+02;
Matches 54; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

```

```

1 aaagtaacacacataatttgataaatttattactaaacacttttctagtaactgttaac 60
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
378 AACGCAATATATATACATACAAATTACTATTTRAAGTTTATTTCAGGCAATTCCTATG 437

QY 61 atgtctgaggattttacagtaataaagaa 89
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 438 TGTCTTGAGGCTGCTTAGGAATTTAGAA 466

RESULT 15
TA195A02P/c TA195A02P 549 bp DNA GSS 13-DEC-2000
LOCUS T. brucei sheared genomic DNA clone 195a02, forward sequence,
DEFINITION genomic survey sequence.
ACCESSION AL477905
VERSION AL477905.1 GI:11841717
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei
REFERENCE Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
AUTHORS Trypanosoma.
1 (bases 1 to 549)
Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,
Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,
Melville,S.E., Rajandream,M.A. and Barrell,B.G.
Direct Submission
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
nh@sanger.ac.uk
COMMENT Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
to give a tight size distribution (
4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).
Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/projects/T_brucei/.
FEATURES
    source
        1..549
            /organism="Trypanosoma brucei"
            /strain="TREU927"
            /db_xref="taxon:5691"
            /clone="195a02"
BASE COUNT 226 a 66 c 97 g 160 t
ORIGIN

Query Match 35.5%; Score 33; DB 13; Length 549;
Best Local Similarity 60.7%; Pred. No. 3.6e+02;
Matches 54; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

```

```

2 aaagtaacacacataatttgataaatttattactaaacacttttctagtaactgttaac 61
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
340 AAATATATATATTTAGGAATATCTAAAAAAGCAAAATTTTCATATTAATTGATAAAA 281

QY 62 tgtctgaggattttacagtaataaagaa 90
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 280 TTTCTTAAATGTTTACCAACTCATAGAA 252

Search completed: April 3, 2002, 04:39:16
Job time: 2836 sec

```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2002, 04:30:10 ; Search time 189.71 Seconds
(without alignments)
54.230 Million cell updates/sec

Title: us-09-700-187-1

Perfect score: 12

Sequence: 1 ggattttacagt 12

Scoring table: IDENTITY_NDC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_1101.*

1:	/SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT.*
2:	/SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT.*
3:	/SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT.*
4:	/SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT.*
5:	/SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT.*
6:	/SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT.*
7:	/SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT.*
8:	/SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT.*
9:	/SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT.*
10:	/SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT.*
11:	/SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT.*
12:	/SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT.*
13:	/SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT.*
14:	/SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT.*
15:	/SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT.*
16:	/SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT.*
17:	/SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT.*
18:	/SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT.*
19:	/SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT.*
20:	/SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT.*
21:	/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT.*
22:	/SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	

1	12	100.0	12	21	AAA97393	
c	2	12	100.0	22	20	AA979146
	3	12	100.0	25	21	AAA68434
	4	12	100.0	30	21	AAA97393
	5	12	100.0	31	21	AAA97411
c	6	12	100.0	31	21	AAA97412
	7	12	100.0	31	21	AAA97423
	8	12	100.0	31	21	AAA97424
	9	12	100.0	31	21	AAA97427
10	12	100.0	31	21	AAA97428	
11	12	100.0	31	22	AA130154	
					Pea pra2 gene high	
					Primer NGA249-R fo	
					Bacteriophage 3A O	
					Pea pra2 light-rep	
					Pea wild-type pra2	
					Pea wild-type pra2	
					Pea wild-type pra2	
					Pea pra2 gene prom	
					Pea pra2 gene prom	
					Pea pra2 gene prom	
					Human single nucle	

c	12	12	100.0	39	21	AAA97403	Pea pra2 light-rep
	13	12	100.0	48	21	AAA97396	Pea pra2 light-rep
	14	12	100.0	52	21	AAA97397	Pea pra2 light-rep
c	15	12	100.0	53	21	AAA97418	Pea wild-type pra2
	16	12	100.0	54	21	AAA97417	Pea wild-type pra2
	17	12	100.0	93	21	AAA97384	Pea pra2 gene ligh
c	18	12	100.0	147	16	AA122540	Human gene signatu
c	19	12	100.0	300	21	AA01517	Human colon cancer
c	20	12	100.0	344	20	AAV90257	EST clone DG279.
c	21	12	100.0	468	22	AA112374	Probe #2307 for ge
c	22	12	100.0	468	22	AA102286	Probe #2277 used t
c	23	12	100.0	547	21	AA93808	Cat flea hindgut a
c	24	12	100.0	646	21	AA256749	Human transmembran
	25	12	100.0	651	22	AAH06265	Human cDNA clone (
	26	12	100.0	724	22	AAH71534	Human cervical can
	27	12	100.0	727	21	AAAC4649	Arabidopsis thalia
	28	12	100.0	852	22	AAH73365	Human cervical can
c	29	12	100.0	1126	21	AAAC4247	Arabidopsis thalia
c	30	12	100.0	1129	20	AA27340	Human secreted pro
c	31	12	100.0	1138	19	AAV46314	Human secreted pro
c	32	12	100.0	1138	22	AA98480	Human cDNA clone B
c	33	12	100.0	1188	20	AAZ52916	Human prostate tum
	34	12	100.0	1221	18	AAV74319	Staphylococcus aur
	35	12	100.0	1279	21	AAAC34384	Arabidopsis thalia
c	36	12	100.0	1287	20	AA37458	Human secreted pro
	37	12	100.0	1341	22	AAAC87638	Arabidopsis thalia
c	38	12	100.0	1342	22	AAH29817	S cerevisiae apopt
c	39	12	100.0	1434	20	AAAX16007	DNA encoding an an
c	40	12	100.0	1466	21	AAAC40161	Arabidopsis thalia
	41	12	100.0	1505	21	AAAC33544	Arabidopsis thalia
c	42	12	100.0	1605	21	AAAC39336	Arabidopsis thalia
c	43	12	100.0	1666	21	AAAC77904	Human cancer assoc
	44	12	100.0	1677	22	AAH34524	Human colon cancer
	45	12	100.0	1706	22	AA98716	Human late stage o

ALIGNMENTS

RESULT 1
AAA97383
ID AAA97383 standard; DNA; 12 BP.
XX
AC AAA97383;
XX
DT 29-JAN-2001 (first entry)
XX
DE Pea pra2 gene light-repressible promoter core region.
XX
KW GTP-binding protein pra2; pea; light-repressible promoter;
KW photoinhibitory; expression cassette; transgenic plant;
KW deterioration prevention; storage; core region; ds.
XX
OS Pisum sativum.
XX
PN WO200055313-AL.
XX
PD 21-SEP-2000.
XX
PF 03-MAR-2000; 2000WO-JP01269.
XX
PR 12-MAR-1999; 99JP-0066551.
XX
(SUNR) SUNTORY LTD.
XX
PI Sasaki Y, Nagano Y, Inaba T;
XX
DR WPI; 2000-587526/55.
XX
PT New DNA fragment or promoter for expressing a target gene, specifically
PT under photoinhibitory conditions, and for transforming a plant cell or
PT plant to improve quality and prevent deterioration during storage -

PS Claim 1; Page 17; 49pp; Japanese.

XX The invention relates to a light-repressible promoter (AA97385), or
CC active fragments thereof (AA97383, AA97384), from the gene encoding
CC the pea Grp-binding protein pra2. The invention also relates to an
CC expression cassette containing the pra2 promoter or its active
CC fragments for the expression of a gene under photoinhibitory or dark
CC conditions in a plant, and to transgenic plants, their descendants
CC and plant tissues comprising the expression cassette. The expression
CC cassette of the invention can be used to generate transgenic plants in
CC which deterioration during storage in the dark is prevented. This is
CC particularly useful for agricultural products. The present sequence
CC represents the pea pra2 promoter core region.

XX Sequence 12 BP; 3 A; 1 C; 3 G; 5 T; 0 other;

Query Match 100.0%; Score 12; DB 21; Length 12;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggattttacagt 12
Db 1 ggattttacagt 12
|||||

RESULT 2
AA979146/C
ID AAX79146 standard; DNA; 22 BP.

XX AAX79146;

DT 17-AUG-1999 (first entry)

DE Primer NGA249-R for A.thaliana SSCP marker.

XX MSH6; MutS homologue; plant; DNA mismatch repair; genetic variation;
KW characteristic; microsatellite; primer; PCR; amplification; SSCP; ss;
KW simple sequence length polymorphism.

XX Synthetic.

OS Arabidopsis thaliana.

XX WO9919492-A2.

PN 22-APR-1999.

XX 09-OCT-1998; 98WO-EP06977.

XX 10-OCT-1997; 97AU-0009745.

PA (RHON) RHONE-POULENC AGROCHIMIE.

PI Betzner AS, Doutriaux M, Freyssinet G, Perez P;

XX WPI; 1999-277644/23.

PT DNA encoding protein functionally involved in the DNA mismatch
PT repair system of a plant

XX Example 3; Page 28; 117pp; English.

XX The invention relates to the isolation of the Arabidopsis thaliana MSH3
CC (AA979066) and MSH6 (AA979067) genes. These genes are MutS homologues
CC (MSH) from plants and are involved in DNA mismatch repair. The DNA
CC sequence can be used in processes for at least partially inactivating a
CC DNA mismatch repair system of a plant, for increasing genetic variation
CC in a plant, and for obtaining a plant with a desired characteristic.
CC Primers AAX79105-X79160 represent 28 primer pairs used to amplify short
CC allelic repeat fragments designated simple sequence length polymorphisms
CC (SSLP). These fragments can be used as markers in the analysis of
CC homologous recombination between genomes of A.thaliana subspecies.

XX

SQ Sequence 22 BP; 7 A; 7 C; 3 G; 5 T; 0 other;

Query Match 100.0%; Score 12; DB 20; Length 22;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggattttacagt 12
Db 21 GGATTTTACAGT 10
|||||

RESULT 3

AAA68434
ID AAA68434 standard; DNA; 25 BP.

XX AAA68434;

XX 27-OCT-2000 (first entry)

DE Bacteriophage 3A ORF RBS sequence 3AORF180.

XX Bacteriophage; antimicrobial; genome; identification; antibacterial;
KW bacterial growth inhibition; PCR primer; RBS; ribosome binding site;
KW bacterial infection; ss.

XX Bacteriophage 3A.

PN WO200032825-A2.

XX 08-JUN-2000.

XX 03-DEC-1999; 99WO-IB02040.

XX 03-DEC-1998; 98US-0110992.

XX 28-SEP-1999; 99US-0326144.

XX 30-SEP-1999; 99US-0407804.

XX 01-DEC-1999; 99US-0157218.

XX 02-DEC-1999; 99US-0168777.

XX (PHAG-) PHAGETECH INC.

XX Pelletier J, Gros P, Dubow M;

XX WPI; 2000-412361/35.

PT Identifying a bacteriophage coding region for treating bacterial
PT infections comprises identifying a nucleic acid encoding a product that
PT inhibits bacteria when a bacteriophage infects a bacterium

XX Disclosure; Page 186; 456pp; English.

XX The present invention describes a method for identifying a bacteriophage
CC coding region encoding a product active on an essential bacterial
CC target. The method comprises identifying a nucleic acid sequence encoding
CC a gene product that provides a bacteria-inhibiting function when an
CC uncharacterised bacteriophage infects a pathogenic bacterium. The
CC compound active on a target of a bacteriophage inhibitor protein in a
CC bacteria is used to treat or prevent a bacterial infection in an animal.
CC AAA68243 to AAA69442 and AAB16523 to AAB16954 represent bacteriophage
CC nucleotide and protein sequences which are used in the exemplification of
CC the present invention.

XX Sequence 25 BP; 5 A; 4 C; 5 G; 11 T; 0 other;

Query Match 100.0%; Score 12; DB 21; Length 25;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggattttacagt 12
|||||


```

PF 03-MAR-2000; 2000WO-JP01269.
XX
PR 12-MAR-1999; 99JP-0066551.
XX
PA (SUNR ) SUNTORY LTD.
XX
XX Sasaki Y, Nagano Y, Inaba T;
XX WPI; 2000-587526/55.
XX
PT New DNA fragment or promoter for expressing a target gene, specifically
PT under photoinhibitory conditions, and for transforming a plant cell or
PT plant to improve quality and prevent deterioration during storage -
XX
PS Example 8; Page 18; 49pp; Japanese.
XX
CC The invention relates to a light-repressible promoter (AAA97385), or
CC active fragments thereof (AAA97383, AAA97384), from the gene encoding
CC the pea GTP-binding protein pra2. The invention also relates to an
CC expression cassette containing the pra2 promoter or its active
CC fragments for the expression of a gene under photoinhibitory or dark
CC conditions in a plant, and to transgenic plants, their descendants
CC and plant tissues comprising the expression cassette. The expression
CC cassette of the invention can be used to generate transgenic plants in
CC which deterioration during storage in the dark is prevented. This is
CC particularly useful for agricultural products. Sequences AAA97411-A97412
CC represent oligonucleotides used in an exemplification of the invention
CC to generate a wild-type pea pra2 promoter fragment.
XX
SQ Sequence 31 BP; 8 A; 7 C; 3 G; 13 T; 0 other;

Query Match 100.0%; Score 12; DB 21; Length 31;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
Db 25 GGATTTTACAGT 14

RESULT 7
AAA97423
ID AAA97423 standard; DNA; 31 BP.
XX
AC AAA97423;
XX
XX 29-JAN-2001 (first entry)
XX
DE Pea wild-type pra2 gene promoter region, bases -672 to -642.
XX
KW GTP-binding protein pra2; pea; light-repressible promoter;
KW photoinhibitory; expression cassette; transgenic plant;
KW deterioration prevention; storage; core region; ds.
XX
OS Pisum sativum.
OS Synthetic.
XX
PN WO200055313-A1.
XX
XX 21-SEP-2000.
XX
XX 03-MAR-2000; 2000WO-JP01269.
XX
XX 12-MAR-1999; 99JP-0066551.
XX
XX (SUNR ) SUNTORY LTD.
XX
XX Sasaki Y, Nagano Y, Inaba T;
XX WPI; 2000-587526/55.
XX
PT New DNA fragment or promoter for expressing a target gene, specifically
PT under photoinhibitory conditions, and for transforming a plant cell or
PT plant to improve quality and prevent deterioration during storage -
XX
PS Example 7; Fig 6a; 49pp; Japanese.
XX
CC The invention relates to a light-repressible promoter (AAA97385), or
CC active fragments thereof (AAA97383, AAA97384), from the gene encoding
CC the pea GTP-binding protein pra2. The invention also relates to an
CC expression cassette containing the pra2 promoter or its active
CC fragments for the expression of a gene under photoinhibitory or dark
CC conditions in a plant, and to transgenic plants, their descendants
CC and plant tissues comprising the expression cassette. The expression
CC cassette of the invention can be used to generate transgenic plants in
CC which deterioration during storage in the dark is prevented. This is
CC particularly useful for agricultural products. The present sequence
CC represents a wild-type pea pra2 promoter fragment (bases -672 to -642 of
CC the pra2 gene) which comprises the pra2 promoter core region (AAA97383).
XX
SQ Sequence 31 BP; 13 A; 3 C; 7 G; 8 T; 0 other;

Query Match 100.0%; Score 12; DB 21; Length 31;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
Db 7 ggattttacagt 18

RESULT 8
AAA97424
ID AAA97424 standard; DNA; 31 BP.
XX
AC AAA97424;
XX
XX 29-JAN-2001 (first entry)
XX
DE Pea pra2 gene promoter region mutant Ls1, bases -672 to -642.
XX
KW GTP-binding protein pra2; pea; light-repressible promoter;
KW photoinhibitory; expression cassette; transgenic plant;
KW deterioration prevention; storage; core region; mutant; ds.
XX
OS Pisum sativum.
OS Synthetic.
XX
PN WO200055313-A1.
XX
XX 21-SEP-2000.
XX
XX 03-MAR-2000; 2000WO-JP01269.
XX
XX 12-MAR-1999; 99JP-0066551.
XX
XX (SUNR ) SUNTORY LTD.
XX
XX Sasaki Y, Nagano Y, Inaba T;
XX WPI; 2000-587526/55.
XX
PT New DNA fragment or promoter for expressing a target gene, specifically
PT under photoinhibitory conditions, and for transforming a plant cell or
PT plant to improve quality and prevent deterioration during storage -
XX
PS Example 7; Fig 6a; 49pp; Japanese.
XX
CC The invention relates to a light-repressible promoter (AAA97385), or
CC active fragments thereof (AAA97383, AAA97384), from the gene encoding
CC the pea GTP-binding protein pra2. The invention also relates to an
CC expression cassette containing the pra2 promoter or its active
CC fragments for the expression of a gene under photoinhibitory or dark
CC conditions in a plant, and to transgenic plants, their descendants
CC and plant tissues comprising the expression cassette. The expression
CC cassette of the invention can be used to generate transgenic plants in
CC which deterioration during storage in the dark is prevented. This is
CC particularly useful for agricultural products. The present sequence
CC represents a wild-type pea pra2 promoter fragment (bases -672 to -642 of
CC the pra2 gene) which comprises the pra2 promoter core region (AAA97383).
XX
SQ Sequence 31 BP; 13 A; 3 C; 7 G; 8 T; 0 other;

```

CC and plant tissues comprising the expression cassette. The expression
 CC cassette of the invention can be used to generate transgenic plants in
 CC which deterioration during storage in the dark is prevented. This is
 CC particularly useful for agricultural products. Sequences AAA97424-A97429
 CC represent mutant pea pra2 promoter fragments (corresponding to bases
 CC -672 to -642 of the gene) which comprise the pra2 promoter core region.
 XX
 SQ Sequence 31 BP; 13 A; 4 C; 7 G; 7 T; 0 other;

Query Match 100.0%; Score 12; DB 21; Length 31;
 Best Local Similarity 100.0%; Pred. No. 4.le+02;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggattttacagt 12
 |||||
 Db 7 ggattttacagt 18

RESULT 9
 AAA97427
 ID AAA97427 standard; DNA; 31 BP.
 XX
 AC AAA97427;
 XX
 DT 29-JAN-2001 (first entry)
 XX
 DE Pea pra2 gene promoter region mutant LS4, bases -672 to -642.
 XX

KW GTP-binding protein pra2; pea; light-repressible promoter;
 KW photoinhibitory; expression cassette; transgenic plant;
 KW deterioration prevention; storage; core region; mutant; ds.
 XX

OS Pisum sativum.
 OS Synthetic.

PN WO200055313-A1.

XX 21-SEP-2000.

XX 03-MAR-2000; 2000WO-JP01269.

XX 12-MAR-1999; 99JP-0066551.

XX (SUNR) SUNTORY LTD.

XX Sasaki Y, Nagano Y, Inaba T;

XX WPI; 2000-587526/55.

XX New DNA fragment or promoter for expressing a target gene, specifically
 XX under photoinhibitory conditions, and for transforming a plant cell or
 XX plant to improve quality and prevent deterioration during storage -
 XX

PS Example 7; Fig 6a; 49pp; Japanese.

XX The invention relates to a light-repressible promoter (AAA97385), or
 CC active fragments thereof (AAA97383, AAA97384), from the gene encoding
 CC the pea GTP-binding protein pra2. The invention also relates to an
 CC expression cassette containing the pra2 promoter or its active
 CC fragments for the expression of a gene under photoinhibitory or dark
 CC conditions in a plant, and to transgenic plants, their descendants
 CC and plant tissues comprising the expression cassette. The expression
 CC cassette of the invention can be used to generate transgenic plants in
 CC which deterioration during storage in the dark is prevented. This is
 CC particularly useful for agricultural products. Sequences AAA97424-A97429
 CC represent mutant pea pra2 promoter fragments (corresponding to bases
 CC -672 to -642 of the gene) which comprise the pra2 promoter core region.
 XX

SQ Sequence 31 BP; 9 A; 5 C; 9 G; 8 T; 0 other;

Query Match 100.0%; Score 12; DB 21; Length 31;

Best Local Similarity 100.0%; Pred. No. 4.le+02;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggattttacagt 12
 |||||
 Db 7 ggattttacagt 18

RESULT 10
 AAA97428
 ID AAA97428 standard; DNA; 31 BP.
 XX
 AC AAA97428;
 XX
 DT 29-JAN-2001 (first entry)
 XX
 DE Pea pra2 gene promoter region mutant LS5, bases -672 to -642.
 XX

KW GTP-binding protein pra2; pea; light-repressible promoter;
 KW photoinhibitory; expression cassette; transgenic plant;
 KW deterioration prevention; storage; core region; mutant; ds.
 XX

OS Pisum sativum.
 OS Synthetic.

PN WO200055313-A1.

XX 21-SEP-2000.

XX 03-MAR-2000; 2000WO-JP01269.

XX 12-MAR-1999; 99JP-0066551.

XX (SUNR) SUNTORY LTD.

XX Sasaki Y, Nagano Y, Inaba T;

XX WPI; 2000-587526/55.

XX New DNA fragment or promoter for expressing a target gene, specifically
 XX under photoinhibitory conditions, and for transforming a plant cell or
 XX plant to improve quality and prevent deterioration during storage -
 XX

PS Example 7; Fig 6a; 49pp; Japanese.

XX The invention relates to a light-repressible promoter (AAA97385), or
 CC active fragments thereof (AAA97383, AAA97384), from the gene encoding
 CC the pea GTP-binding protein pra2. The invention also relates to an
 CC expression cassette containing the pra2 promoter or its active
 CC fragments for the expression of a gene under photoinhibitory or dark
 CC conditions in a plant, and to transgenic plants, their descendants
 CC and plant tissues comprising the expression cassette. The expression
 CC cassette of the invention can be used to generate transgenic plants in
 CC which deterioration during storage in the dark is prevented. This is
 CC particularly useful for agricultural products. Sequences AAA97424-A97429
 CC represent mutant pea pra2 promoter fragments (corresponding to bases
 CC -672 to -642 of the gene) which comprise the pra2 promoter core region.
 XX

SQ Sequence 31 BP; 11 A; 4 C; 7 G; 9 T; 0 other;

Query Match 100.0%; Score 12; DB 21; Length 31;
 Best Local Similarity 100.0%; Pred. No. 4.le+02;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggattttacagt 12
 |||||
 Db 7 ggattttacagt 18

RESULT 11
 AA130154
 ID AA130154 standard; DNA; 31 BP.

```

XX AC AAI30154;
XX DT 18-OCT-2001 (first entry)
XX DE Human single nucleotide polymorphism (SNP) PCCB 3.
XX KW Human; resequence; genotype; disease; forensic; paternity testing;
XX KW single nucleotide polymorphism; SNP; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Variation replace(16,A)
XX FT /*tag= a
XX FT /standard_name= "single nucleotide polymorphism"
XX PN WO200166800-A2.
XX PD 13-SEP-2001.
XX PF 07-MAR-2001; 2001WO-US07268.
XX PR 07-MAR-2000; 2000US-0187510.
XX PR 22-MAY-2000; 2000US-0206129.
XX PA (WHED ) WHITEHEAD INST BIOMEDICAL RES.
XX PI Cargill M, Ireland JS, Lander ES;
XX DR WPI; 2001-522952/57.
XX PT Nucleic acid molecules from the human genome which include polymorphic
XX PT sites, useful in methods for predicting the presence, absence or
XX PT severity of a particular phenotype or disorder (e.g. diabetes)
XX PT associated with a particular genotype
XX PS Claim 1; Page 69; 145pp; English.
XX CC The invention relates to the identification of nucleic acid molecules
XX CC (AAI29513-AAI31314) from the human genome which include polymorphic sites
XX CC which can predispose individuals to disease. Various genes from a number
XX CC of individuals were resequenced and single nucleotide polymorphisms
XX CC (SNPs) in these genes discovered. The method is useful for predicting the
XX CC presence, absence or severity of a particular phenotype or disorder (e.g.
XX CC diabetes) associated with a particular genotype. The nucleic acids
XX CC containing the polymorphic sites may be useful in forensics and paternity
XX CC testing.
XX SE Sequence 31 BP; 6 A; 4 C; 5 G; 16 T; 0 other;

Query Match 100.0%; Score 12; DB 22; Length 31;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
DB 16 ggattttacagt 27

RESULT 12
AAI97403/C
ID AAA97403 standard; DNA; 39 BP.
XX AC AAA97403;
XX DT 29-JAN-2001 (first entry)
XX DE Pea pra2 light-repressible promoter clone PL4C PCR primer, SEQ ID NO:21.
XX KW GTP-binding protein pra2; pea; light-repressible promoter;
XX KW photoinhibitory; expression cassette; transgenic plant;

Query Match 100.0%; Score 12; DB 21; Length 39;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
DB 37 GGATTTTACAGT 26

RESULT 13
AAI97396
ID AAA97396 standard; DNA; 48 BP.
XX AC AAA97396;
XX DT 29-JAN-2001 (first entry)
XX DE Pea pra2 light-repressible promoter mutant LS4 PCR primer, SEQ ID NO:14.
XX KW GTP-binding protein pra2; pea; light-repressible promoter;
XX KW photoinhibitory; expression cassette; transgenic plant;
XX KW deterioration prevention; storage; mutagenic PCR primer; ss.
XX OS Pisum sativum.
XX OS Synthetic.
XX PN WO200055313-A1.
XX PD 21-SEP-2000.
XX PF 03-MAR-2000; 2000WO-JP01269.
XX PR 12-MAR-1999; 99JP-0066551.

```

```

KW deterioration prevention; storage; PCR primer; ss.
XX Pisum sativum.
XX WO200055313-A1.
XX PD 21-SEP-2000.
XX PF 03-MAR-2000; 2000WO-JP01269.
XX PR 12-MAR-1999; 99JP-0066551.
XX PA (SUNR ) SUNTORY LTD.
XX PI Sasaki Y, Nagano Y, Inaba T;
XX DR WPI; 2000-587526/55.
XX PT New DNA fragment or promoter for expressing a target gene, specifically
XX PT under photoinhibitory conditions, and for transforming a plant cell or
XX PT plant to improve quality and prevent deterioration during storage
XX PS Example 3; Page 14; 49pp; Japanese.
XX CC The invention relates to a light-repressible promoter (AAA97385), or
XX CC active fragments thereof (AAA97383, AAA97384), from the gene encoding
XX CC the pea GTP-binding protein pra2. The invention also relates to an
XX CC expression cassette containing the pra2 promoter or its active
XX CC fragments for the expression of a gene under photoinhibitory or dark
XX CC conditions in a plant, and to transgenic plants, their descendants
XX CC and plant tissues comprising the expression cassette. The expression
XX CC cassette of the invention can be used to generate transgenic plants in
XX CC which deterioration during storage in the dark is prevented. This is
XX CC particularly useful for agricultural products. Sequences AAA97387-A97392
XX CC and AAA97398-A97410 represent PCR primers used in an exemplification of
XX CC the invention to amplify the pea pra2 promoter for cloning.
XX SE Sequence 39 BP; 7 A; 8 C; 7 G; 17 T; 0 other;

Query Match 100.0%; Score 12; DB 21; Length 39;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
DB 37 GGATTTTACAGT 26

RESULT 13
AAI97396
ID AAA97396 standard; DNA; 48 BP.
XX AC AAA97396;
XX DT 29-JAN-2001 (first entry)
XX DE Pea pra2 light-repressible promoter mutant LS4 PCR primer, SEQ ID NO:14.
XX KW GTP-binding protein pra2; pea; light-repressible promoter;
XX KW photoinhibitory; expression cassette; transgenic plant;
XX KW deterioration prevention; storage; mutagenic PCR primer; ss.
XX OS Pisum sativum.
XX OS Synthetic.
XX PN WO200055313-A1.
XX PD 21-SEP-2000.
XX PF 03-MAR-2000; 2000WO-JP01269.
XX PR 12-MAR-1999; 99JP-0066551.

```

```

XX PA (SUNR ) SUNTORY LTD.
XX PI
XX PI Sasaki Y, Nagano Y, Inaba T;
XX DR WPI; 2000-587526/55.
XX
XX PT New DNA fragment or promoter for expressing a target gene, specifically
XX PT under photoinhibitory conditions, and for transforming a plant cell or
XX PT plant to improve quality and prevent deterioration during storage -
XX PS
XX PS Example 3; Page 13; 49pp; Japanese.
XX CC The invention relates to a light-repressible promoter (AAA97385), or
XX CC active fragments thereof (AAA97383, AAA97384), from the gene encoding
XX CC the pea GTP-binding protein pra2. The invention also relates to an
XX CC expression cassette containing the pra2 promoter or its active
XX CC fragments for the expression of a gene under photoinhibitory or dark
XX CC conditions in a plant, and to transgenic plants, their descendants
XX CC and plant tissues comprising the expression cassette. The expression
XX CC cassette of the invention can be used to generate transgenic plants in
XX CC which deterioration during storage in the dark is prevented. This is
XX CC particularly useful for agricultural products. Sequences AAA97393-A97398
XX CC represent PCR primers used in an exemplification of the invention
XX CC to mutate the pea pra2 promoter.
XX SQ Sequence 52 BP; 19 A; 9 C; 12 G; 12 T; 0 other;

Query Match 100.0%; Score 12; DB 21; Length 52;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
Db | | | | | | | | | |
13 ggattttacagt 24

RESULT 15
AAA97418/c
ID AAA97418 standard; DNA; 53 BP.
XX AC AAA97418;
XX DT 29-JAN-2001 (first entry)
XX DE Pea wild-type pra2 gene light-repressible promoter oligonucleotide, WT4.
XX KW GTP-binding protein pra2; pea; light-repressible promoter;
XX KW photoinhibitory; expression cassette; transgenic plant;
XX KW deterioration prevention; storage; ss.
XX OS Pisum sativum.
XX PN WO200055313-A1.
XX PD 21-SEP-2000.
XX PF 03-MAR-2000; 2000WO-JP01269.
XX PR 12-MAR-1999; 99JP-0066551.
XX PA (SUNR ) SUNTORY LTD.
XX PI Sasaki Y, Nagano Y, Inaba T;
XX PI WPI; 2000-587526/55.
XX DR
XX DR New DNA fragment or promoter for expressing a target gene, specifically
XX DR under photoinhibitory conditions, and for transforming a plant cell or
XX DR plant to improve quality and prevent deterioration during storage -
XX PS
XX PS Example 3; Page 19; 49pp; Japanese.
XX CC The invention relates to a light-repressible promoter (AAA97385), or
XX CC active fragments thereof (AAA97383, AAA97384), from the gene encoding
XX CC the pea GTP-binding protein pra2. The invention also relates to an
XX CC expression cassette containing the pra2 promoter or its active
XX CC fragments for the expression of a gene under photoinhibitory or dark
XX CC conditions in a plant, and to transgenic plants, their descendants
XX CC and plant tissues comprising the expression cassette. The expression
XX CC cassette of the invention can be used to generate transgenic plants in
XX CC which deterioration during storage in the dark is prevented. This is
XX CC particularly useful for agricultural products. Sequences AAA97417-A97418
XX CC represent PCR primers used in an exemplification of the invention
XX CC to mutate the pea pra2 promoter.
XX SQ Sequence 48 BP; 15 A; 9 C; 13 G; 11 T; 0 other;

Query Match 100.0%; Score 12; DB 21; Length 48;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
Db | | | | | | | | | |
13 ggattttacagt 24

RESULT 14
AAA97397
ID AAA97397 standard; DNA; 52 BP.
XX AG AAA97397;
XX DT 29-JAN-2001 (first entry)
XX DE Pea pra2 light-repressible promoter mutant LS5 PCR primer, SEQ ID NO:15.
XX KW GTP-binding protein pra2; pea; light-repressible promoter;
XX KW photoinhibitory; expression cassette; transgenic plant;
XX KW deterioration prevention; storage; mutagenic PCR primer; ss.
XX OS Pisum sativum.
XX OS Synthetic.
XX PN WO200055313-A1.
XX PD 21-SEP-2000.
XX PF 03-MAR-2000; 2000WO-JP01269.
XX PR 12-MAR-1999; 99JP-0066551.
XX PA (SUNR ) SUNTORY LTD.
XX PI Sasaki Y, Nagano Y, Inaba T;
XX PI WPI; 2000-587526/55.
XX DR
XX DR New DNA fragment or promoter for expressing a target gene, specifically
XX DR under photoinhibitory conditions, and for transforming a plant cell or
XX DR plant to improve quality and prevent deterioration during storage -
XX PS
XX PS Example 3; Page 13; 49pp; Japanese.
XX CC The invention relates to a light-repressible promoter (AAA97385), or
XX CC active fragments thereof (AAA97383, AAA97384), from the gene encoding
XX CC the pea GTP-binding protein pra2. The invention also relates to an
XX CC expression cassette containing the pra2 promoter or its active
XX CC fragments for the expression of a gene under photoinhibitory or dark
XX CC conditions in a plant, and to transgenic plants, their descendants
XX CC and plant tissues comprising the expression cassette. The expression
XX CC cassette of the invention can be used to generate transgenic plants in
XX CC which deterioration during storage in the dark is prevented. This is
XX CC particularly useful for agricultural products. Sequences AAA97393-A97398
XX CC represent PCR primers used in an exemplification of the invention
XX CC to mutate the pea pra2 promoter.
XX SQ Sequence 52 BP; 19 A; 9 C; 12 G; 12 T; 0 other;

```

CC represent oligonucleotides used in an exemplification of the invention
CC to generate a wild-type pea pr2 promoter fragment.

XX
SQ Sequence 53 BP; 21 A; 11 C; 3 G; 18 T; 0 other;

Query Match 100.0%; Score 12; DB 21; Length 53;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggattttacagt 12
|||||
Db 33 GGATTTTACAGT 22

Search completed: April 3, 2002, 05:18:30
Job time: 2900 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2002, 05:13:00 ; Search time 1999.71 Seconds
(without alignments)
767.230 Million cell updates/sec

Title: US-09-700-187-2

Perfect score: 93

Sequence: 1 aaagtaacacatatatttga.....ttacagtaataaagaacga 93

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_cm:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_om:*
- 20: em_or:*
- 21: em_ov:*
- 22: em_pat:*
- 23: em_ph:*
- 24: em_pl:*
- 25: em_ro:*
- 26: em_sts:*
- 27: em_sy:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_htgo_hum:*
- 31: em_htgo_inv:*
- 32: em_htgo_rod:*
- 33: em_htg_hum:*
- 34: em_htg_inv:*
- 35: em_htg_rod:*
- 36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Query Match	Length	DB	ID	Description
1	93	100.0	3441	8	AB007911	AB007911 Pisum sat
2	37.6	40.4	157381	9	AL360267	AL360267 Human DNA
3	37.2	40.0	2498	1	BTNRNA	U03552 Bacillus th
4	36.4	39.1	340	3	AFU77847	U77847 Aconophora
5	36.4	39.1	133021	9	AL161624	AL161624 Human DNA
6	36.4	39.1	169671	2	AC021098	AC021098 Homo sapi
7	35.6	38.3	141279	2	AC023824	AC023824 Homo sapi
8	35.6	38.3	142654	9	AC012049	AC012049 Homo sapi
9	35.6	38.3	154937	9	AC022537	AC022537 Homo sapi
10	35.4	38.1	50143	2	AC027071	AC027071 Homo sapi
11	35.4	38.1	175691	2	AC007615	AC007615 Homo sapi
12	35.4	38.1	191496	2	AC009130	AC009130 Homo sapi
13	35.4	38.1	210043	2	AC009093	AC009093 Homo sapi
14	35	37.6	721	8	AF072528	AF072528 Lycopersi
15	35	37.6	35836	9	AL160414	AL160414 Human DNA
16	34.6	37.2	646	8	AF365230	AF365230 Tetraberl
17	34.6	37.2	66084	8	AB015468	AB015468 Arabidops
18	34.4	37.0	148249	2	AL451133	AL451133 Homo sapi
19	34.4	37.0	157604	9	AC021093	AC021093 Homo sapi
20	34.4	37.0	189938	2	AL355588	AL355588 Homo sapi
21	34.4	37.0	193729	2	AC023772	AC023772 Homo sapi
22	34.2	36.8	293181	1	AP001119	AP001119 Buchnera
23	34	36.6	43152	3	CELF59B1	AF067943 Caenorhab
24	34	36.6	84194	8	AB006701	AB006701 Arabidops
25	34	36.6	104992	2	AC005504	AC005504 Plasmodiu
26	34	36.6	130513	9	AC004592	AC004592 Homo sapi
27	34	36.6	136060	2	AC004153	AC004153 Plasmodiu
28	34	36.6	169546	2	AC004157	AC004157 Plasmodiu
29	34	36.6	245805	2	AC006752	AC006752 Caenorhab
30	34	36.6	300193	3	AE003426	AE003426 Drosophil
31	33.8	36.3	74227	9	HSJ493H23	AL121789 Human DNA
32	33.8	36.3	148018	2	AP003974	AP003974 Oryza sat
33	33.8	36.3	184604	9	AL138682	AL138682 Human DNA
34	33.8	36.3	174617	2	AP003977	AP003977 Oryza sat
35	33.6	36.1	143549	2	AP003910	AP003910 Oryza sat
36	33.6	36.1	168030	2	AC091462	AC091462 Mus muscu
37	33.2	35.7	37219	9	AC091515	AC091515 Homo sapi
38	33.2	35.7	188296	2	AC022324	AC022324 Homo sapi
39	33.2	35.7	188470	2	AC093005	AC093005 Homo sapi
40	33	35.5	656	8	AF365229	AF365229 Tetraberl
41	33	35.5	3707	2	AC012967	AC012967 Drosophil
42	33	35.5	58407	1	MI12CG	L71118 Methanococc
43	33	35.5	135599	8	CPU30821	U30821 Cyanophora
44	33	35.5	139064	2	AC073440	AC073440 Homo sapi
45	33	35.5	153549	2	AC022664	AC022664 Homo sapi

ALIGNMENTS

RESULT 1
AB007911
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AB007911 3441 bp DNA PLN
Pisum sativum gene for PRA2, complete cds. 28-SEP-1999

AB007911
AB007911.1 GI:5926717
PRA2.

Pisum sativum DNA.

Pisum sativum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae;
Pisum.

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE

1 (sites)
Inaba,T., Nagano,Y., Sakakibara,T. and Sasaki,Y.
Identification of a cis-regulatory element involved in phytochrome
down-regulated expression of the pea small GTPase gene praz
Plant Physiol. 120 (2), 491-500 (1999)
99292944
2 (bases 1 to 3441)

AUTHORS Nagano,Y.
 TITLE Direct Submission
 JOURNAL Submitted (07-OCT-1997) Yukio Nagano, Nagoya University, Graduate School of Bioagricultural Sciences: Chikusa, Nagoya, Aichi 464-8601, Japan (E-mail:nagano@agr.nagoya-u.ac.jp, Tel:81-52-789-4168, Fax:81-52-789-4296)

FEATURES
 source
 1..3441
 /organism="Pisum sativum"
 /db_xref="taxon:3888"
 repeat_unit 904..1018
 gene join(2326..2567,2729..3188)
 /gene="pra2"
 CDS join(2326..2567,2729..3188)
 /gene="pra2"
 /note="light-repressible GTP binding protein"
 /codon_start=1
 /product="PRA2"
 /protein_id="BAA84640.1"
 /db_xref="GI:5926718"
 /translation="MNQEMNGVEAEKLOEKIDYFVKVVGDSAVGKTOILSRFTKNE
 CFQSKSTIGVEFOTKVTYINGKLKAIQIDTAGOERYRAVTSAYVRCALGAMLVYDI
 TKRQTFDHAARVVEELRSHADGSIIVMLIGNKGLDVGQVQTEDAVEFAEDQGLFFS
 ETSAPSGENVASAFLLKLOEKINKVVKRSLECNKNGKNGNDHVAALAGEKIDIIISAS
 ELEISEIKKLHSCSC"

BASE COUNT 1143 a 606 c 540 g 1152 t
 ORIGIN

Query Match 100.0%; Score 93; DB 8; Length 3441;
 Best Local Similarity 100.0%; Pred. No. 1.3e-10;
 Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aaagtaacacatatattgataaattattactaaactattttctagctactgttaac 60
 |||||
 Db 1396 AAAGTAACACATATTTGATGAATTTATTAATAAATTTCTTAGTACTGTTAATC 1455
 |||||

QY 61 atgtctgaggattttacagtaataaagaacga 93
 |||||
 Db 1456 ATGTCTGAGGATTTTACAGTAATAAAGAACGA 1488
 |||||

RESULT 2
 AL360267/c 157381 bp DNA PRI 22-MAY-2001
 LOCUS Human DNA sequence from clone RP11-342C20 on chromosome 13,
 DEFINITION complete sequence.
 ACCESSION AL360267
 VERSION AL360267.10 GI:13751441
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 157381)
 Tracey,A.
 Direct Submission
 Submitted (22-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On Apr 21, 2001 this sequence version replaced gi:13443424.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the

assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em., EMBL; Sw.,
 SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP
 database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 13, constructed by the Sanger Centre Chromosome 13
 Mapping Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr13
 RP11-342C20 is from the library RPI-11.2 constructed by the group
 of Pieter de Jong. For further details see
 http://www.chori.org/bacpac/home.htm
 VECTOR: pBAC3.6
 This sequence is the entire insert of clone RP11-342C20 The true
 left end of clone RP11-365J7 is at 99075 in this sequence.

FEATURES
 source
 1..157381
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="13"
 /clone="RP11-342C20"
 /clone_lib="RPI-11.2"
 misc_feature 64251..64259
 /note="Sequence from uni-directional dGTP big dye
 terminator reads only."
 121811..121816
 /note="Sequence from overlapping clone BA365J7 (AL359472).
 Assembly confirmed by restriction digest."
 BASE COUNT 51039 a 27179 c 27761 g 51402 t
 ORIGIN

Query Match 40.4%; Score 37.6; DB 9; Length 157381;
 Best Local Similarity 63.0%; Pred. No. 27;
 Matches 58; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 2 aaagtaacacatatattgataaattattactaaactattttctagctactgttaac 61
 |||||
 Db 42229 AAATAACAGACATTTTGATCATTTGTAATATAAGCTAGATAGTAAATGCTACAAA 42170
 |||||

QY 62 tgctcgaggattttacagtaataaagaacga 93
 |||||
 Db 42169 TATTTGAAATCTTTTCATTAGTCAAGCAACA 42138
 |||||

RESULT 3
 BTHKNA 2498 bp DNA BCT 01-APR-1995
 LOCUS Bacillus thuringiensis histidine protein kinase (hknA) gene,
 DEFINITION complete cds.
 ACCESSION U03552
 VERSION U03552.1 GI:495313
 KEYWORDS
 SOURCE Bacillus thuringiensis.
 ORGANISM Bacillus thuringiensis
 Bacteria; Firmicutes; Bacillus/Clostridium group;
 Bacillus/Staphylococcus group; Bacillus; Bacillus cereus group.
 1 (bases 1 to 2498)
 Baum,J.A.
 Tn340J, a new class II transposable element from Bacillus
 thuringiensis
 J. Bacteriol. 176 (10), 2835-2845 (1994)
 JOURNAL
 MEDLINE 94245608
 REFERENCE 2 (bases 1 to 2498)
 AUTHORS Baum,J.A.
 TITLE Direct Submission
 Submitted (17-NOV-1993) Jim A. Baum, Ecogen Inc., Strain
 Development, Molecular Genetics, 2005 Cabot Boulevard West,
 Langhorne, PA 19047-1810, USA
 JOURNAL
 FEATURES Location/Qualifiers
 source 1..2498
 /organism="Bacillus thuringiensis"


```
/strain="EG7566"
/sub_species="kurstaki"
/db_xref="taxon:1428"
682..1800
/gene="hkna"
682..1800
/gene="hkna"
/standard_name="Hkna"
/codon_start=1
/transl_table=11
/function="phosphorylation"
/evidence-experimental
/product="histidine protein kinase"
/protein_id="AA04586.1"
/db_xref="GI:520402"
/transl_table="MEVFPDKDIKEIFCSHLKNNRHOFVNNKMKMIISEKDPFKLE
VWONGDLELLEIEMEDKDINVLQCEKIALERAGADANIGDFVNNVNGRNELF
EAMCDELVSARELKPINAKIHTCFDKLIYTVLKYSILSKNLEEKQOYINETHKREL
TIIGQMSASVHEFRNPLTSIMGFVKLLKADHPSLSYLDIISHELDQLNKRSIOFLV
SKEMNESERFWLNDLFDIIQFLPSLVNANVLIENKLPYIPILVGYSEVRQVEL
NILMNSIDALESKEERKIIIDVFEDQAIRIVIKNGPMIPAENVETIPEPFTVTKK
LGTGIGLVCKQIVKEHNGSIMCRSDNDWTEFQIAFOK"
BASE COUNT      858 a 329 c 444 g 867 t
ORIGIN

Query Match      40.0%; Score 37.2; DB 1; Length 2498;
Best Local Similarity 63.3%; Pred. No. 56;
Matches 57; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

Qy 1 aaagtaacacatatattgtataaaattattactaaactattttctagctactgttaac 60
||||| ||| ||||| ||||| ||| ||| ||||| ||||| |||
Db 1042 AAATACATACACTGTTTGACAAATTAATTATTATTATACCGTTTTAAATACTCGAAATT 1101
||||| ||| ||||| ||||| ||| ||| ||||| ||||| |||

Qy 61 atgtctgaggattttacagtaataaagaaa 90
||||| ||| ||||| ||||| ||| ||| ||||| ||||| |||
Db 1102 ATATCGAAGAATTTAGAGGAAAAACAGCAA 1131
||||| ||| ||||| ||||| ||| ||| ||||| ||||| |||

RESULT 4
AFU77847 LOCUS      340 bp DNA INV 23-DEC-1996
DEFINITION Aconophora ferruginea 12S mitochondrial ribosomal RNA, small
subunit, mitochondrial gene, partial sequence.
ACCESSION U77847
VERSION U77847.1 GI:1750156
KEYWORDS Aconophora ferruginea.
SOURCE Mitochondrion Aconophora ferruginea
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Paraneoptera; Hemiptera; Euhemiptera;
Membracidae; Membracidae; Aconophora.
REFERENCE 1 (bases 1 to 340)
AUTHORS Liu,D.
TITLE Direct Submission
JOURNAL Submitted (12-NOV-1996) Exploratory Research, Human Genome
Sciences, 9410 Key West Ave., Rockville, MD 20850, USA
FEATURES
source 1..340
/organism="Aconophora ferruginea"
/organella="mitochondrion"
<1..>340
/product="12S small subunit ribosomal RNA"
BASE COUNT      131 a 25 c 43 g 141 t
ORIGIN

Query Match      39.1%; Score 36.4; DB 3; Length 340;
Best Local Similarity 66.7%; Pred. No. 1.le+02;
Matches 52; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Qy 12 atatttgataaatttactaaactattttctagctactgttaacatctgctgagga 71
```

```
||||| ||| ||||| ||||| ||| ||| ||||| ||||| |||
Db 228 AAATATGGTTTACTTTATTAGTAATAATTATATAACTAGATGTTTAAATTTAGGA 287
||||| ||| ||||| ||||| ||| ||| ||||| ||||| |||
Qy 72 ttttacagtaataaagaa 89
||||| ||| ||||| ||||| ||| ||| ||||| ||||| |||
Db 288 TTTAAAGTAATATAAATA 305

RESULT 5
AL161624/c LOCUS      133021 bp DNA PRI 18-MAY-2001
DEFINITION Human DNA sequence from clone RP11-48719 on chromosome X, complete
sequence.
ACCESSION AL161624
VERSION AL161624.7 GI:14160916
KEYWORDS HTG.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 133021)
AUTHORS Bird,C.
TITLE Direct Submission
JOURNAL Submitted (18-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
COMMENT On May 20, 2001 this sequence version replaced gi:13274323.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, EMBL; Sw,
SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome X, constructed by the Sanger Centre Chromosome X Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/ChrX
RP11-48719 is from the library RPCI-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-48719 it may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone RP11-212B22 is at 132922 in this
sequence. The true right end of clone RP11-274M8 is at 100 in this
sequence.
FEATURES
source 1..133021
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/clone="RP11-48719"
/clone_lib="RPCI-11.2"
118..425
/note="Alusq repeat: matches 1..300 of consensus"
982..1042
/note="MER69 repeat: matches 1..62 of consensus"
1311..1622
/note="AluSp repeat: matches 1..310 of consensus"
2450..2552
repeat_region
repeat_region
repeat_region
repeat_region
```



```

QY      64  ttgaggattttacagta 81
      || | | | | | | |
DB 77389  ACTCACTAATTATTATATA 77372

RESULT 6
AC021098/c
LOCUS      AC021098.3   169671 bp   DNA           HTG           01-SEP-2000
DEFINITION Homo sapiens chromosome X clone RP11-33A2, WORKING DRAFT SEQUENCE,
            18 unordered pieces.
ACCESSION  AC021098
VERSION    AC021098.3   GI:7230836
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 169671)
            Waterston,R.H.
            The sequence of Homo sapiens clone
            2 (bases 1 to 169671)
            Waterston,R.H.
            Direct Submission
            Submitted (14-JAN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
            On Mar 13, 2000 this sequence version replaced gi:6922906.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0033A02
----- Summary Statistics -----
Sequencing vector: M13; 81%
Sequencing vector: plasmid; 19%
Chemistry: Dye-primer ET; 81% of reads
Chemistry: Dye-terminator Big Dye; 19% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 159126 bases at least Q40
Consensus quality: 162659 bases at least Q30
Consensus quality: 164792 bases at least Q20
Insert size: 170000; agarose-fp
Insert size: 167971; sum-of-contigs
Quality coverage: 4.16 in Q20 bases; agarose-fp
Quality coverage: 4.21 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1196: contig of 1196 bp in length
* 1197: gap of unknown length
* 1297: contig of 1988 bp in length
* 3285: gap of unknown length
* 3385: contig of 2639 bp in length
* 6024: gap of unknown length
* 6124: gap of unknown length
* 8768: contig of 2645 bp in length
* 8769: gap of unknown length
* 8869: contig of 3303 bp in length
* 12172: gap of unknown length
* 12272: contig of 3706 bp in length
* 15978: gap of unknown length
* 16078: contig of 8173 bp in length
* 24251: gap of unknown length
* 32681: contig of 8331 bp in length
* 32682: gap of unknown length

```

```

* 32782 41939: contig of 9158 bp in length
* 41940 42039: gap of unknown length
* 42040 48802: contig of 6763 bp in length
* 48803 48902: gap of unknown length
* 48903 57427: contig of 8525 bp in length
* 57428 57527: gap of unknown length
* 57528 66423: contig of 8796 bp in length
* 66424 78283: gap of unknown length
* 78284 78384: gap of unknown length
* 78384 93662: contig of 15279 bp in length
* 93663 93762: gap of unknown length
* 93763 111527: contig of 17764 bp in length
* 111527 129714: gap of unknown length
* 129715 129814: contig of 18088 bp in length
* 129815 144537: contig of 14723 bp in length
* 144538 14638: gap of unknown length
* 14638 169671: contig of 25034 bp in length.

FEATURES             Location/Qualifiers
     source            1..169671
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /clone="RP11-33A2"
     misc_feature      1..1196
                     /note="assembly_name:Contig7"
     misc_feature      1297..3284
                     /note="assembly_name:Contig8"
     misc_feature      3385..6023
                     /note="assembly_name:Contig9"
     misc_feature      6124..8768
                     /note="assembly_name:Contig10"
     misc_feature      8869..12171
                     /note="assembly_name:Contig11"
     misc_feature      12272..15977
                     /note="assembly_name:Contig12"
     misc_feature      16078..24250
                     /note="assembly_name:Contig13"
     misc_feature      24351..32681
                     /note="assembly_name:Contig14"
     misc_feature      32782..41939
                     /note="assembly_name:Contig15"
     misc_feature      42040..48802
                     /note="assembly_name:Contig16"
                     clone_end:SP6
                     vector_side:right
     misc_feature      48903..57427
                     /note="assembly_name:Contig17"
                     clone_end:T7
                     vector_side:right
     misc_feature      57528..66423
                     /note="assembly_name:Contig18"
     misc_feature      66424..78283
                     /note="assembly_name:Contig19"
     misc_feature      78384..93662
                     /note="assembly_name:Contig20"
     misc_feature      93763..111526
                     /note="assembly_name:Contig21"
     misc_feature      111627..129714
                     /note="assembly_name:Contig22"
     misc_feature      129815..144537
                     /note="assembly_name:Contig23"
     misc_feature      144638..169671
                     /note="assembly_name:Contig24"

BASE COUNT  53919 a 31234 c 31915 g 50898 t 1705 others
ORIGIN

```

```

Query Match          39.1%; Score 36.4; DB 2; Length 169671;
Best Local Similarity 66.7%; Pred. NO. 48;
Matches 52; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

```



```

BASE COUNT      51071 a 29071 c 26776 g 48019 t
ORIGIN

Query Match      38.3%; Score 35.6; DB 9; Length 154937;
Best Local Similarity 62.2%; Pred. No. 71;
Matches 56; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 1 aaagtagcacacatttttgatataatttattactaaaactattttctagctactgttaac 60
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 87539 AAAAAATTAACCCCATGTTACTTTTTCAGTAGTACTAAAAAAGTACTTTTAAATTTATGTCATC 87598

QY 61 atgtctgagagattttacagtaataaagaaa 90
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 87599 ATGTTAGTAACTACTAAAAACAAGAAAGAAA 87628

RESULT 10
AC027071/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AC027071 50143 bp DNA HTG 27-MAR-2000
Homo sapiens chromosome 9 clone RP11-200D20 map 9, LOW-PASS
SEQUENCE SAMPLING.
AC027071
AC027071.1 GI:7330331
HTG; HTGS_PHASE0.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 50143)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bedalov,F.,
Boguski,M., Bouckge,J., Brown,A., Burkett,G.,
Campiano,A., Castle,A., Choehel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (27-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit,A.F.P. & Green,P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WITR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L7662
Center clone name: 200_D_20
-----
* NOTE: This record contains 57 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads

```



```

RESULT 11
AC007615      175691 bp      DNA      HTG      03-JUL-2001
LOCUS      Homo sapiens chromosome 16 clone RP11-528K16, WORKING DRAFT
DEFINITION      SEQUENCE, 1 ordered pieces.
AC007615
AC007615      175691 bp      GI:14589428
VERSION      HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
KEYWORDS      human.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 175691)
AUTHORS      DOE Joint Genome Institute.
TITLE      Sequencing of Human Chromosome 16
JOURNAL      Unpublished
AUTHORS      2 (bases 1 to 175691)
Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,
Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,
Bryant,J., Tesmer,J., Meincke,L., Longmire,J., White,S., Tatam,O.,
Campbell,C., Fawcett,J., Maltbie,M., Bussod,M., Sutherland,R.,
McMurry,K., Han,C. and Deaven,L.
Direct Submission
Submitted (20-MAY-1999) Center for Human Genome Studies, DOE Joint
Genome Institute, Los Alamos National Laboratory, MS M888, Los
Alamos, NM 87545, USA
COMMENT      Sequence Quality Assessment:
On Jul 3, 2001 this sequence version replaced gi:13928651.
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.
-----
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 175691: contig of 175691 bp in length.
FEATURES
    source
        1..175691
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="16"
            /clone="RP11-528K16"
BASE COUNT      42392 a 42125 c 45277 g 45896 t      1 others
ORIGIN
Query Match      38.1%; Score 35.4; DB 2; Length 175691;
Best Local Similarity 66.2%; Pred. No. 77;
Matches 51; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
QY      2 aaagtaacacatattttgataaattattactaaactattttctagttactgttaataca 61
      | ||||| ||||||||| ||||||||| ||||| ||| | ||||| |||

```

```

Db 97786 ATAGTAGCTCATATTGTTAAATTTATCCCTAAGAATTTTATGATTCACTTGCAAATGG 97845
QY      62 tqtctgaggagttttaca 78
      | | | | | ||||| | |
Db 97846 TATTTTATTTTTCGCA 97862

RESULT 12
AC009130/c      191496 bp      DNA      HTG      21-JUN-2000
LOCUS      Homo sapiens chromosome 16 clone RP11-501P17, WORKING DRAFT
DEFINITION      SEQUENCE, 43 unordered pieces.
AC009130
AC009130.6      GI:8575965
VERSION      HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS      human.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 191496)
AUTHORS      DOE Joint Genome Institute.
TITLE      Sequencing of Human Chromosome 16
JOURNAL      Unpublished
AUTHORS      2 (bases 1 to 191496)
DOE Joint Genome Institute.
Direct Submission
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jun 21, 2000 this sequence version replaced gi:7689948.
COMMENT      -----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 599126
Center clone name: RPCI-11_501P17
-----
Summary Statistics
Consensus quality: 158339 bases at least Q40
Consensus quality: 176831 bases at least Q30
Consensus quality: 180633 bases at least Q20
Estimated insert size: 201840; agarose-fp estimation
Estimated insert size: 187296; sum-of-contigs estimation
Quality coverage: 3.4 in Q20 bases; agarose-fp estimation
Quality coverage: 3.67 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 43 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1071: contig of 1071 bp in length
* 1072 1171: gap of unknown length
* 1172 2358: contig of 1187 bp in length
* 2359 2458: gap of unknown length
* 2459 3680: contig of 1222 bp in length
* 3681 3780: gap of unknown length
* 3781 5278: contig of 1498 bp in length
* 5279 5379: gap of unknown length
* 5379 6522: contig of 1144 bp in length
* 6523 6622: gap of unknown length
* 6623 8589: contig of 1967 bp in length
* 8590 8690: gap of unknown length
* 8690 10266: contig of 1577 bp in length
* 10267 10366: gap of unknown length
* 10367 11714: contig of 1348 bp in length
* 11715 11814: gap of unknown length
* 11815 13443: contig of 1629 bp in length
* 13444 13543: gap of unknown length
* 13544 14645: contig of 1102 bp in length

```


* 8718 8817: gap of unknown length
 * 8818 10104: contig of 1287 bp in length
 * 10105 10204: gap of unknown length
 * 10205 12294: contig of 2090 bp in length
 * 12295 12394: gap of unknown length
 * 12395 15488: contig of 3094 bp in length
 * 15489 18845: gap of unknown length
 * 18846 18945: gap of unknown length
 * 18946 23168: contig of 4223 bp in length
 * 23169 23268: gap of unknown length
 * 23269 29524: contig of 6256 bp in length
 * 29525 29624: gap of unknown length
 * 29625 54542: contig of 24918 bp in length
 * 54543 54642: gap of unknown length
 * 54643 81600: contig of 26958 bp in length
 * 81601 81700: gap of unknown length
 * 81701 115005: contig of 33305 bp in length
 * 115006 115105: gap of unknown length
 * 115106 161572: contig of 46467 bp in length
 * 161573 161672: gap of unknown length
 * 161673 210043: contig of 48371 bp in length.

FEATURES

source
 1. .210043
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="16"
 /clone="RP11-426C22"
 /clone_lib="RPC1 human BAC library 11"
 53123 a 52905 c 51736 g 50777 t 1502 others

BASE COUNT
 ORIGIN

Query Match 38.1%; Score 35.4; DB 2; Length 210043;
 Best Local Similarity 66.2%; Pred. No. 75;
 Matches 51; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
 QY 2 aaagtaacacatatttgataaaatttattactaaacattttctagctactgttaataca 61
 Db 110362 ATAGTACCTATATTTGTTAAATTTATCCCTAAGAAATTTATGATTCACCTGCAATGG 110303
 QY 62 tgtctgaggattttaca 78
 Db 110302 TATTTATTTTGTGCA 110286

RESULT 14
 AF072528 721 bp DNA PLN 29-SEP-1998
 LOCUS
 DEFINITION Lycopersicon pennellii clone AG15 paracentromeric sequence.
 ACCESSION AF072528
 VERSION AF072528.1 GI:3342056
 KEYWORDS
 SOURCE Lycopersicon pennellii.
 ORGANISM Lycopersicon pennellii
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
 Lycopersicon.
 REFERENCE 1 (bases 1 to 721)
 AUTHORS Weide,R., Hontelez,J., van Kammen,A., Koornneef,M. and Zabel,P.
 TITLE Paracentromeric sequences on tomato chromosome 6 show homology to human satellite III and to the mammalian CENP-B binding box
 JOURNAL Mol. Gen. Genet. 259 (2), 190-197 (1998)
 MEDLINE 98418482
 REFERENCE 2 (bases 1 to 721)
 AUTHORS Weide,R., Hontelez,J., van Kammen,A., Koornneef,M. and Zabel,P.
 TITLE Direct Submission
 JOURNAL Submitted (19-JUN-1998) Phytopathology, Wageningen Agricultural University, Binnenhaven 9, Wageningen 6709 PD, the Netherlands
 FEATURES Location/Qualifiers
 1. .721
 /organism="Lycopersicon pennellii"

/strain="LA 716"
 /db_xref="taxon:28526"
 /chromosome="6"
 /map="centromere"
 /clone="AG15"
 239 a 69 c 116 g 297 t
 BASE COUNT
 ORIGIN

Query Match 37.6%; Score 35; DB 8; Length 721;
 Best Local Similarity 61.5%; Pred. No. 1.9e+02;
 Matches 56; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 3 aagtaacacatatttgataaaatttattactaaacattttctagctactgttaataca 62
 Db 326 AAATAACCCATAGTTATTTATATATTTAAGTAAAGATTGTCCAGATCATTTTAATTTT 385

QY 63 gtctgaggattttacagtaataaagaacga 93
 Db 386 TAATATTAGTTTATATATGATAATGTAACATA 416

RESULT 15
 AL160414/c
 LOCUS
 DEFINITION

AL160414 35836 bp DNA PRI 15-FEB-2001
 Human DNA sequence from clone RP5-931h19 on chromosome 20. Contains ESTs, GSSs and CpG islands. Contains the OXT gene for oxytocin (neurophysin I) and the AVP gene for arginine vasopressin (neurophysin II), complete sequence.
 AL160414
 ACCESSION AL160414.18 GI:9863596
 VERSION
 KEYWORDS HTG; AVP; CpG island; neurophysin; OXT; oxytocin; vasopressin.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 35836)
 AUTHORS Heath,P.
 TITLE Direct Submission
 JOURNAL Submitted (13-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 COMMENT requests: clonerequest@sanger.ac.uk
 On Aug 21, 2000 this sequence version replaced gi:9843543.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:
 Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep
 This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr20
 IMPORTANT: This sequence is not the entire insert of clone RP5-931h19 it may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
 The true left end of clone RP5-1187M17 is at 35737 in this sequence. The true right end of clone RP4-534B8 is at 100 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP5-931h19 is from the library RPC1-5 constructed by the group of Pieter de Jong. For further details see
 http://www.chori.org/bacpac/home.htm

FEATURES	VECTOR: pCYPAC2.		
source	Location/Qualifiers		
	1..35836	repeat_region	/note="HERV16 repeat: matches 2908..3069 of consensus"
	/organism="Homo sapiens"	repeat_region	9412..9705
	/db_xref="taxon:9606"	repeat_region	/note="AluSx repeat: matches 16..310 of consensus"
	/chromosome="20"	repeat_region	9717..10018
	/clone="RP5-931H19"	repeat_region	/note="AluSg repeat: matches 1..306 of consensus"
	/clone_lib="RPC1-5"	repeat_region	10020..10072
	1..158	repeat_region	/note="L1MA9 repeat: matches 6259..6308 of consensus"
repeat_region	/note="AluY repeat: matches 1..158 of consensus"	repeat_region	10218..10818
repeat_region	2142..2408	repeat_region	/note="L1M2 repeat: matches 95..425 of consensus"
repeat_region	/note="AluSx repeat: matches 1..296 of consensus"	repeat_region	10827..10929
repeat_region	2834..3122	repeat_region	/note="L1M1 repeat: matches 2202..2684 of consensus"
repeat_region	/note="AluSx repeat: matches 1..290 of consensus"	repeat_region	10993..11252
repeat_region	3133..3363	repeat_region	/note="AluSg repeat: matches 1..301 of consensus"
repeat_region	/note="AluSx repeat: matches 1..231 of consensus"	repeat_region	11320..12087
repeat_region	3557..3870	repeat_region	/note="L1 repeat: matches 3006..3748 of consensus"
repeat_region	/note="AluDb repeat: matches 1..307 of consensus"	repeat_region	12191..12246
repeat_region	4187..4417	repeat_region	/note="L1P5 repeat: matches 4436..4491 of consensus"
repeat_region	/note="MIR repeat: matches 40..262 of consensus"	repeat_region	12247..12552
repeat_region	4419..4723	repeat_region	/note="AluSg repeat: matches 3..306 of consensus"
repeat_region	/note="AluSp repeat: matches 1..312 of consensus"	repeat_region	12553..13393
repeat_region	4954..5248	repeat_region	/note="L1P5 repeat: matches 4491..5324 of consensus"
repeat_region	/note="AluSp repeat: matches 3..298 of consensus"	repeat_region	13394..13696
repeat_region	5282..5593	repeat_region	/note="AluSg repeat: matches 1..304 of consensus"
repeat_region	/note="AluSx repeat: matches 1..310 of consensus"	repeat_region	13697..13739
repeat_region	5594..5704	repeat_region	/note="L1P5 repeat: matches 5324..5366 of consensus"
repeat_region	/note="L1MC4 repeat: matches 7701..7813 of consensus"	repeat_region	13741..13842
repeat_region	5766..5904	repeat_region	/note="AluJo/FRAM repeat: matches 198..299 of consensus"
repeat_region	/note="FLAM_C repeat: matches 1..133 of consensus"	repeat_region	13843..14144
repeat_region	5905..6242	repeat_region	/note="AluY repeat: matches 1..299 of consensus"
repeat_region	/note="L1MA5A repeat: matches 5478..5827 of consensus"	repeat_region	14149..14687
repeat_region	6243..6546	repeat_region	/note="SVA repeat: matches 8..548 of consensus"
repeat_region	/note="AluSx repeat: matches 18..309 of consensus"	repeat_region	14692..14799
repeat_region	6631..6679	repeat_region	/note="L1P5 repeat: matches 5315..5425 of consensus"
misc_feature	/note="L1M4 repeat: matches 2698..2746 of consensus"	repeat_region	14815..15106
misc_feature	6701..7023	repeat_region	/note="AluSx repeat: matches 1..294 of consensus"
misc_feature	/note="match: GSS: Em:AQ318472"	repeat_region	15108..15271
repeat_region	6701..7051	repeat_region	/note="AluDb repeat: matches 133..307 of consensus"
repeat_region	/note="match: GSS: Em:AQ782226"	repeat_region	15272..15296
repeat_region	6721..7033	repeat_region	/note="L1M1 repeat: matches 5807..5828 of consensus"
repeat_region	/note="AluSp repeat: matches 1..313 of consensus"	repeat_region	15297..16003
repeat_region	complement(7032..7218)	repeat_region	/note="L1T8 repeat: matches 1..691 of consensus"
repeat_region	/note="match: GSS: Em:AQ236587 Em:AQ239242"	repeat_region	16004..16110
repeat_region	complement(7033..7172)	repeat_region	/note="L1M1 repeat: matches 5828..5952 of consensus"
repeat_region	/note="match: GSS: Em:AQ423172"	repeat_region	16191..16264
repeat_region	complement(7034..7218)	repeat_region	/note="L1M3 repeat: matches 6056..6129 of consensus"
repeat_region	/note="match: GSS: Em:AQ286166 Em:AQ596537"	repeat_region	16324..16396
repeat_region	complement(7037..7219)	repeat_region	/note="L2 repeat: matches 1370..1454 of consensus"
repeat_region	/note="match: GSS: Em:AQ614522"	repeat_region	16436..16508
repeat_region	complement(7045..7218)	repeat_region	/note="L2 repeat: matches 1582..1649 of consensus"
repeat_region	7048..7220	repeat_region	16509..16811
repeat_region	/note="match: GSS: Em:AQ545857"	repeat_region	/note="AluSp repeat: matches 1..306 of consensus"
repeat_region	7233..7297	repeat_region	16812..17029
repeat_region	/note="L1M1 repeat: matches 1190..1252 of consensus"	repeat_region	/note="L2 repeat: matches 1649..1896 of consensus"
repeat_region	complement(join(7233..7296,7617..7774))	repeat_region	17030..17337
repeat_region	/note="match: GSS: Em:AQ541198"	repeat_region	/note="AluSx repeat: matches 1..307 of consensus"
repeat_region	7298..7614	repeat_region	17346..17497
repeat_region	/note="AluJo repeat: matches 6..307 of consensus"	repeat_region	/note="AluSg/x repeat: matches 135..288 of consensus"
repeat_region	7615..7798	repeat_region	17499..17688
repeat_region	/note="L1M1 repeat: matches 998..1190 of consensus"	repeat_region	/note="L2 repeat: matches 1881..2104 of consensus"
repeat_region	7800..7932	repeat_region	17689..17983
repeat_region	/note="AluSg/x repeat: matches 1..133 of consensus"	repeat_region	/note="AluY repeat: matches 3..297 of consensus"
repeat_region	7937..8074	repeat_region	17984..18023
repeat_region	/note="AluSg repeat: matches 1..140 of consensus"	repeat_region	/note="L2 repeat: matches 2104..2138 of consensus"
repeat_region	8075..8382	repeat_region	18126..18218
repeat_region	/note="AluA5 repeat: matches 5..311 of consensus"	repeat_region	/note="L2 repeat: matches 2303..2398 of consensus"
repeat_region	8383..8550	repeat_region	18346..18628
repeat_region	/note="AluSg repeat: matches 140..302 of consensus"	repeat_region	/note="AluSx repeat: matches 1..299 of consensus"
repeat_region	8730..9040	repeat_region	18629..18694
repeat_region	/note="AluDb repeat: matches 2..312 of consensus"	repeat_region	/note="Alu repeat: matches 246..296 of consensus"
repeat_region	9060..9207	repeat_region	18944..19131
		repeat_region	/note="L1TIE repeat: matches 295..538 of consensus"
		repeat_region	19132..19439
		repeat_region	/note="AluJo repeat: matches 1..305 of consensus"

```
repeat_region 19440..19473
/note="MLTIE repeat: matches 538..564 of consensus"
repeat_region 19506..19744
/note="L2 repeat: matches 2465..2685 of consensus"
repeat_region 20242..20618

Query Match      37.6%; Score 35; DB 9; Length 35836;
Best Local Similarity 70.1%; Pred. No. 1.1e+02;
Matches 47; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Qy 13 tatttgataaattattactaaactattttcttagtactgtgtaataatcatgctgaggat 72
    ||||| || ||||| ||||| ||||| || ||||| || ||||| || ||||| || |||||
Db 14715 TATTTTATTTATTTATTTTAAAAAAATTTTATAGTATTTTATTCATCATTCTTGGGTGT 14656

Qy 73 ttacag 79
    || || ||
Db 14655 TTCTCGG 14649
```

Search completed: April 3, 2002, 05:14:06
Job time: 4161 sec

THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2002, 05:14:51 ; Search time 81.95 seconds
(without alignments)
257.016 Million cell updates/sec

Title: US-09-700-187-2
Perfect score: 93
Sequence: 1 aaagtaacacatatattga.....ttacagtaataaagaacga 93

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 11323999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2.6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2.6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2.6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2.6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2.6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2.6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	29.4	31.6	1984	1	US-07-885-970A-25
C 2	29.4	31.6	1985	1	US-08-298-687A-25
C 3	29.4	31.6	1985	1	US-08-298-829-25
C 4	29.4	31.6	6373	3	US-08-462-728-1
C 5	29.4	31.6	6375	1	US-08-168-917-5
C 6	29.4	31.6	6375	2	US-08-460-510-5
C 7	29.4	31.6	6375	2	US-08-460-490-5
C 8	29.4	31.6	6375	5	PCT-US92-00730-5
C 9	29.4	31.6	6375	5	PCT-US92-00862-5
C 10	28.2	30.3	392	4	US-09-385-982-95
C 11	28.2	30.3	4576	1	US-08-832-883-49
C 12	28.2	30.3	4576	2	US-08-832-877-49
C 13	28.2	30.3	4673	1	US-07-638-431-1
C 14	28.2	30.3	4673	5	PCT-US92-00018-1
C 15	28.2	30.3	6138	4	US-09-067-800-4
C 16	28.2	30.3	6138	4	US-09-349-677-4
C 17	28	28	2290	6	5312912-1
C 18	27.6	29.7	1947	3	US-08-604-991-1
C 19	27.6	29.7	1947	3	US-09-363-639-1
C 20	27.4	29.5	617	4	US-09-385-982-222
C 21	27.4	29.5	967	4	US-08-960-780-47
C 22	27.4	29.5	967	4	US-09-073-898-47
C 23	27.2	29.2	713	4	US-08-998-416-956
C 24	27.2	29.2	3974	4	US-08-467-504-3
C 25	27.2	29.2	5336	4	US-09-102-528-11
C 26	27.2	29.2	6471	4	US-09-353-585-1
C 27	27	29.0	3095	6	5231168-1

C 28	26.8	28.8	1431	4	US-09-316-083-2	Sequence 2, Appli
C 29	26.6	28.6	591	4	US-09-328-111-807	Sequence 807, App
C 30	26.6	28.6	663	4	US-08-998-416-187	Sequence 187, App
C 31	26.6	28.6	696	4	US-08-998-416-779	Sequence 779, App
C 32	26.6	28.6	719	4	US-08-998-416-1138	Sequence 1138, Ap
C 33	26.6	28.6	856	4	US-08-998-416-289	Sequence 289, App
C 34	26.6	28.6	2481	2	US-08-958-642-3	Sequence 3, Appli
C 35	26.6	28.6	2481	3	US-08-778-394-1	Sequence 1, Appli
C 36	26.6	28.6	2481	3	US-08-778-423A-3	Sequence 3, Appli
C 37	26.6	28.6	13158	2	US-08-687-080-105	Sequence 105, App
C 38	26.4	28.4	662	4	US-08-998-416-185	Sequence 185, App
C 39	26.4	28.4	701	4	US-08-998-416-701	Sequence 701, App
C 40	26.4	28.4	724	4	US-08-998-416-683	Sequence 683, App
C 41	26.4	28.4	732	4	US-08-998-416-1036	Sequence 1036, Ap
C 42	26.4	28.4	767	4	US-08-998-416-472	Sequence 472, App
C 43	26.4	28.4	782	4	US-08-998-416-224	Sequence 224, App
C 44	26.4	28.4	827	4	US-08-998-416-535	Sequence 535, App
C 45	26.4	28.4	828	4	US-08-998-416-538	Sequence 538, App

ALIGNMENTS

RESULT 1
US-07-885-970A-25/c
; Sequence 25, Application US/07885970A
; Patent No. 5495070
; GENERAL INFORMATION:
; APPLICANT: John, Maliyakal E.
; TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nicholas J. Seay, Quarles & Brady
; STREET: P.O. Box 2113, First Wisconsin Plaza
; CITY: Madison
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53701

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/885,970A
FILING DATE: 19920518
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/617,239
FILING DATE: 21-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/253,243
FILING DATE: 04-OCT-1988
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27,386
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 283-2478
TELEFAX: (608) 251-5139
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 1984 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Gossypium barbadense
STRAIN: Sea Island
IMMEDIATE SOURCE:
LIBRARY: EMBL S1

CLONE: SIH6
US-07-885-970A-25

	Query Match	31.6%;	Score 29.4;	DB 1;	Length 1984;
	Best Local Similarity	58.6%;	Pred. No. 8.1;		
	Matches 51;	Conservative 0;	Mismatches 36;	Indels 0;	Gaps
QY	1	aaagtaaacacataatttggataaattattactaaacactatttctagctactgtttaac	60		
Db	1023	AAAAATTAAATTTAAATTTAAACCAATTAATAAAATTATATTTAAAGTCTTGCTGTA	984		
QY	61	atgtctgaggattttcacgtaataag	87		
Db	963	AGGCTTAAATTTCAACCGATAATAAAG	937		

```

RESULT      2
US-08-298-687A-25/c
; Sequence 25, Application US/08298687A
; Patent No. 5521078
; GENERAL INFORMATION:
; APPLICANT: John, Malliyakal E.
; TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON
; TITLE OF INVENTION: PLANTS FOR ALTERED FIBER
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nicholas J. Seay, Quarles & Brady
; STREET: P.O. Box 2113, First Wisconsin Plaza
; CITY: Madison
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53701

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/298,687A
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/617,239
FILING DATE: 21-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/253,243
FILING DATE: 04-OCT-1988
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27,386
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 283-2478
TELEFAX: (608) 251-5139
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 1985 base pairs
TYPE: nucleic acid
STRAINEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Gossypium barbadense
STRAIN: Sea Island
IMMEDIATE SOURCE:
LIBRARY: EMBL-SI
CLONE: SIH6
US-08-298-687A-25

Query Match 31.6%; Score 29.4; DB 1; Length 1985;

	Best Local Similarity	58.6%;	Pred. No. 8.1;						
	Matches	51;	Conservative	0;	Mismatches	36;	Indels	0;	Gaps
QY	1	aaagtaaacacataatttggataaattactactaaaactattttctagctacttggtaatc	60						
Db	1023	AAATATATATTTAATATTTATAAAACCATTAATAAATTATATTTTAAGTCTGCTGTA	964						
QY	61	atgtctgaggattttcacagtaataaag	87						
Db	963	AGGCTTAAATTTCAACCGATAATAAAG	937						

RESULT 3
US-08-298-829-25/c
: Sequence 25. Application us/08298829

GENERAL INFORMATION:
 APPLICANT: John, Mallyakal E.
 TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON
 TITLE OF INVENTION: PLANTS FOR ALTERED FIBER
 NUMBER OF SEQUENCES: 33
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Nicholas J. Seay, Quarles & Brady
 STREET: P. O. Box 2113, First Wisconsin Plaza
 CITY: Madison
 STATE: Wisconsin
 COUNTRY: USA
 ZIP: 53701

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM pc compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/298,829
FILING DATE: 19-OCT-1994
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/885,970
FILING DATE: 18-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/617,239
FILING DATE: 21-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/253,243
FILING DATE: 04-OCT-1988
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27,386
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 283-2478
TELEFAX: (608) 251-5139
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 1985 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Gossypium barbadense
STRAIN: Sea Island
IMMEDIATE SOURCE:
LIBRARY: EMBL-SI
CLONE: STH6
US-08-298-829-25

```

Query Match 31.6%; Score 29.4; DB 1; Length 1985;
Best Local Similarity 58.6%; Pred. No. 8.1;
Matches 51; Conservative 0; Mismatches 36; Indels 0

[illegible]


```
PCT-US92-00730-5
; Sequence 5, Application PC/TUS9200730
; GENERAL INFORMATION:
; APPLICANT: Wolf, David
; APPLICANT: Tomlinson, James E.
; APPLICANT: Fretto, Larry J.
; APPLICANT: Giese, Neill A.
; APPLICANT: Escobedo, Jaime A.
; APPLICANT: Williams, Lewis T.
; TITLE OF INVENTION: DOMAINS OF EXTRACELLULAR REGION OF HUMAN
; TITLE OF INVENTION: PLATELET-DERIVED GROWTH FACTOR RECEPTOR POLYPEPTIDES
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND
; STREET: Steuart Street Tower, 20th Floor \ One Market
; STREET: Plaza
; CITY: San Francisco
; STATE: California
; COUNTRY: US
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/00730
; FILING DATE: 19920128
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P.
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: 12418-14
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 326-2400
; TELEFAX: (415) 326-2422
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6375 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo Sapiens
; STRAIN: lambda gt10
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 129..3395
; OTHER INFORMATION: /note= "nucleotide number 1 of this
; OTHER INFORMATION: sequence is identical to the nucleotide number 1
; OTHER INFORMATION: of the previous 4100 long sequence"
PCT-US92-00730-5

Query Match 31.68; Score 29.4; DB 5; Length 6375;
Best Local Similarity 58.6%; Pred. No. 8.9; Indels 0; Gaps 0;
Matches 51; Conservative 0; Mismatches 36;

Qy 4 agtaacacattttgataaaatttactataaaactattttctagctactgttaactcatg 63
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6029 AGTTATACATATATACATAAAGATATATCTCGAACCTCTTATGACGGTTTTGTAAATACTG 6088

Qy 64 tctgaggattttacagtaataaagaaa 90
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6089 TTCGACATAGTCACGGAAGCAAAATATA 6115

RESULT 9
PCT-US92-00862-5
; Sequence 5, Application PC/TUS9200862
```

```
; GENERAL INFORMATION:
; APPLICANT: Williams, Lewis T.
; APPLICANT: Escobedo, Jaime A.
; TITLE OF INVENTION: PLATELET-DERIVED GROWTH FACTOR RECEPTORS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND
; STREET: Steuart Street Tower, 20th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: US
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/00862
; FILING DATE: 19920131
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/151,141
; FILING DATE: 02-FEB-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/309,322
; FILING DATE: 10-FEB-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P.
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: 23070-267-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 326-2400
; TELEFAX: (415) 326-2422
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6375 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHEtical: N
; ANTI-SENSE: N
; ORIGINAL SOURCE:
; ORGANISM: Homo Sapiens
; IMMEDIATE SOURCE:
; LIBRARY: lambda gt10
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 129..3395
; OTHER INFORMATION: /note= "nucleotide number 1 of this
; OTHER INFORMATION: sequence is identical to the nucleotide
; OTHER INFORMATION: number 1 of the previous 4100 long
; OTHER INFORMATION: sequence."
PCT-US92-00862-5

Query Match 31.68; Score 29.4; DB 5; Length 6375;
Best Local Similarity 58.6%; Pred. No. 8.9; Indels 0; Gaps 0;
Matches 51; Conservative 0; Mismatches 36;

Qy 4 agtaacacattttgataaaatttactataaaactattttctagctactgttaactcatg 63
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6029 AGTTATACATATATACATAAAGATATATCTCGAACCTCTTATGACGGTTTTGTAAATACTG 6088

Qy 64 tctgaggattttacagtaataaagaaa 90
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6089 TTCGACATAGTCACGGAAGCAAAATATA 6115

RESULT 10
US-09-385-982-95
; Sequence 95, Application US/09385982
```

```
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCDA-260XX
; CURRENT APPLICATION NUMBER: US/09/395,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 95
; LENGTH: 392
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)-(392)
; OTHER INFORMATION: n = A,T,C or G
US-09-383-982-95

; Query Match          30.3%; Score 28.2; DB 4; Length 392;
; Best Local Similarity 54.8%; Pred. No. 15;
; Matches 51; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 1 aaagtaacacatttgataaattattactaaactattttctagctactgttaatc 60
    |||| || || || || || || || || || || || || || || || || ||
Db 249 aaaaaaagaagttgtgaattttattacttttaagtgtgatactaaagtattaaac 308
    || || || || || || || || || || || || || || || || || || ||

QY 61 atgtctgaggattttacagtaataaagaacga 93
    || || || || || || || || || || || || || || || || || || ||
Db 309 atattctgnattcttccaaaaaanaanta 341
    || || || || || || || || || || || || || || || || || || ||

RESULT 11
US-08-832-883-49/c
; Sequence 49, Application US/08832883
; Patent No. 5807681
; GENERAL INFORMATION:
; APPLICANT: Giordano, Antonio
; TITLE OF INVENTION: METHODS FOR THE DIAGNOSIS AND PROGNOSIS
; TITLE OF INVENTION: OF CANCER
; NUMBER OF SEQUENCES: 115
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEIDEL, GONDA, LAVORGNA & MONACO, P.C.
; STREET: Suite 1800 Two Penn Center Plaza
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/832,883
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Monaco, Daniel A
; REGISTRATION NUMBER: 30,480
; REFERENCE/DOCKET NUMBER: 8321-13 US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-8383
; TELEFAX: (215) 568-5549

; Query Match          30.3%; Score 28.2; DB 2; Length 4576;
; Best Local Similarity 61.6%; Pred. No. 18;
; Matches 45; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 6 taacacataattttgataaattattactaaaaactattttctagctactgttaatc 65
    || || || || || || || || || || || || || || || || || || ||
Db 3477 TTACGTGTAGTTAAACCACTTTTATTATTATCTTAATTTCTAGAAATTAAGAAATTATTC 3418
    || || || || || || || || || || || || || || || || || || ||

; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4576 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-832-883-49

; Query Match          30.3%; Score 28.2; DB 1; Length 4576;
; Best Local Similarity 61.6%; Pred. No. 18;
; Matches 45; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 6 taacacataattttgataaattattactaaaaactattttctagctactgttaatc 65
    || || || || || || || || || || || || || || || || || || ||
Db 3477 TTACGTGTAGTTAAACCACTTTTATTATTATCTTAATTTCTAGAAATTAAGAAATTATTC 3418
    || || || || || || || || || || || || || || || || || || ||

; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4576 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-832-877-49

; Query Match          30.3%; Score 28.2; DB 2; Length 4576;
; Best Local Similarity 61.6%; Pred. No. 18;
; Matches 45; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 6 taacacataattttgataaattattactaaaaactattttctagctactgttaatc 65
    || || || || || || || || || || || || || || || || || || ||
Db 3477 TTACGTGTAGTTAAACCACTTTTATTATTATCTTAATTTCTAGAAATTAAGAAATTATTC 3418
    || || || || || || || || || || || || || || || || || || ||
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2002, 05:18:30 ; Search time 189.71 Seconds
(without alignments)
420.280 Million cell updates/sec

Title: US-09-700-187-2

Perfect score: 93

Sequence: 1 aaagtaacacattttga.....ttacagtaataaagaacga 93

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	93	100.0	93	21	AA97384
2	93	100.0	2325	21	AA97385
3	93	100.0	3441	21	AA97382
c 4	33.6	36.1	936	22	AAF58252
c 5	33.6	36.1	936	22	AAF58254
c 6	33.6	36.1	936	22	AAF58257
c 7	33.6	36.1	936	22	AAF58259
c 8	33.6	36.1	936	22	AAF58262
c 9	33.6	36.1	938	22	AAF58255
10	33	35.5	58407	19	AAV21210
11	32.8	35.3	244	22	AAF58238

c	12	32.8	35.3	244	22	AAF58238	Oligonucleotide D1
	13	32.8	35.3	936	22	AAF58252	Oligonucleotide D1
	14	32.8	35.3	936	22	AAF58254	Oligonucleotide D1
	15	32.8	35.3	936	22	AAF58257	Oligonucleotide D1
	16	32.8	35.3	936	22	AAF58259	Oligonucleotide D2
	17	32.8	35.3	936	22	AAF58262	Oligonucleotide D2
	18	32.8	35.3	938	22	AAF58255	Oligonucleotide D1
c	19	32.4	34.8	5535	21	AA70184	Plasmodium falciparum
	20	32.2	34.6	269223	22	AAF28554	Genomic fragment #
c	21	31.8	34.2	3549	21	AA70223	Plasmodium falciparum
c	22	31.6	34.0	441	21	AA70237	Human secreted protein
c	23	31	33.3	31	21	AA97411	Pea wild-type pra2
c	24	31	33.3	31	21	AA97412	Pea wild-type pra2
c	25	31	33.3	31	21	AA97423	Pea wild-type pra2
c	26	31	33.3	507	22	AA06274	Domestic mite Bt2
c	27	30.6	32.9	134	16	AA724176	Human gene signatu
c	28	30.2	32.5	1301	21	AA62843	Lipid transfer pro
c	29	30	32.3	1431	21	AA237082	DNA sequence encod
c	30	30	32.3	1671	13	AAQ24134	50 kD subunit of S
c	31	30	32.3	5763	18	AAV74816	Staphylococcus aur
c	32	30	32.3	5924	18	AAV74441	Staphylococcus aur
c	33	29.8	32.0	3001	21	AAH51805	Chromosome 13q31-q
c	34	29.6	31.8	600	18	AA74092	DNA encoding a Sta
c	35	29.4	31.6	877	21	AA54965	Arabidopsis thalia
c	36	29.4	31.6	878	21	AA37095	Arabidopsis thalia
c	37	29.4	31.6	1984	17	AA713030	Cotton fibre-speci
c	38	29.4	31.6	1985	17	AA730250	Cotton fibre clone
c	39	29.4	31.6	1985	18	AA770036	Cotton H6 gene and
c	40	29.4	31.6	2099	21	AA57997	2099 bp Candida al
c	41	29.4	31.6	6075	13	AAQ27451	Type A human plate
c	42	29.4	31.6	129021	21	AA722296	BAC containing rep
c	43	29.4	31.6	580073	18	AA758840	Mycoplasma genital
c	44	29.4	31.6	910715	20	AA720248	Borrelia burgdorfe
c	45	29.2	31.4	475	21	AAH30428	Human colon cancer

ALIGNMENTS

RESULT 1
AA97384
ID AAA97384 standard; DNA; 93 BP.
XX
AC AAA97384;
XX
XX
DT 29-JAN-2001 (first entry)
XX
DE Pea pra2 gene light-repressible promoter cis element.
XX
XX GTP-binding protein pra2; pea; light-repressible promoter;
KW photoinhibitory; expression cassette; transgenic plant;
KW deterioration prevention; storage; cis element; ds.
XX
OS Pisum sativum.
XX
XX WO200055313-A1.
XX
XX 21-SEP-2000.
XX
XX 03-MAR-2000; 2000WO-JP01269.
XX
XX 12-MAR-1999; 99JP-0066551.
XX
XX (SUNR) SUNTORY LTD.
XX
XX Sasaki Y, Nagano Y, Inaba T;
XX
XX WPI; 2000-587526/55.
XX
XX New DNA fragment or promoter for expressing a target gene, specifically
PT under photoinhibitory conditions, and for transforming a plant cell or
PT plant to improve quality and prevent deterioration during storage
XX

PS Claim 2; Page 33; 49pp; Japanese.

XX The invention relates to a light-repressible promoter (AAA97385), or

CC active fragments thereof (AAA97383, AAA97384), from the gene encoding

CC the pea GTP-binding protein pra2. The invention also relates to an

CC expression cassette containing the pra2 promoter or its active

CC fragments for the expression of a gene under photoinhibitory or dark

CC conditions in a plant, and to transgenic plants, their descendants

CC and plant tissues comprising the expression cassette. The expression

CC cassette of the invention can be used to generate transgenic plants in

CC which deterioration during storage in the dark is prevented. This is

CC particularly useful for agricultural products. The present sequence

CC represents the pea pra2 promoter cis element.

XX

SQ Sequence 93 BP; 38 A; 10 C; 11 G; 34 T; 0 other;

Query Match 100.0%; Score 93; DB 21; Length 93;

Best Local Similarity 100.0%; Pred. No. 2.7e-14;

Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aaagtaacacatatatttgataaattattactaaactattttctagctactgttaac 60

|||||

Db 1 aaagtaacacatatatttgataaattattactaaactattttctagctactgttaac 60

|||||

QY 61 atgtctgaggattttacagtaataaagaacga 93

|||||

Db 61 atgtctgaggattttacagtaataaagaacga 93

|||||

RESULT 2

AAA97385

ID AAA97385 standard; DNA; 2325 BP.

XX

AC AAA97385;

XX

DT 29-JAN-2001 (first entry)

XX

DE Pea pra2 gene light-repressible promoter.

XX

KW GTP-binding protein pra2; pea; light-repressible promoter;

KW photoinhibitory; expression cassette; transgenic plant;

KW deterioration prevention; storage; ds.

XX

OS Pisum sativum.

XX

PN WO200055313-A1.

XX

PD 21-SEP-2000.

XX

PF 03-MAR-2000; 2000WO-JP01269.

XX

PR 12-MAR-1999; 99JP-0066551.

XX

PA (SUNR) SUNTORY LTD.

XX

PI Sasaki Y, Nagano Y, Inaba T;

XX

PS WPI; 2000-587526/55.

XX

DR

XX

PT New DNA fragment or promoter for expressing a target gene, specifically

PT under photoinhibitory conditions, and for transforming a plant cell or

PT plant to improve quality and prevent deterioration during storage -

XX

PS Claim 3; Page 34-35; 49pp; Japanese.

XX

CC The invention relates to a light-repressible promoter (AAA97385), or

CC active fragments thereof (AAA97383, AAA97384), from the gene encoding

CC the pea GTP-binding protein pra2. The invention also relates to an

CC expression cassette containing the pra2 promoter or its active

CC fragments for the expression of a gene under photoinhibitory or dark

CC conditions in a plant, and to transgenic plants, their descendants

CC and plant tissues comprising the expression cassette. The expression

CC cassette of the invention can be used to generate transgenic plants in

CC which deterioration during storage in the dark is prevented. This is

CC particularly useful for agricultural products. The present sequence

CC represents the pea pra2 gene.

XX

SQ Sequence 3441 BP; 1143 A; 606 C; 540 G; 1152 T; 0 other;

Query Match 100.0%; Score 93; DB 21; Length 2325;

Best Local Similarity 100.0%; Pred. No. 2.8e-14;

Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aaagtaacacatatatttgataaattattactaaactattttctagctactgttaac 60

|||||

Db 1396 aaagtaacacatatatttgataaattattactaaactattttctagctactgttaac 1455

|||||

QY 61 atgtctgaggattttacagtaataaagaacga 93

|||||

Db 1456 atgtctgaggattttacagtaataaagaacga 1488

|||||

RESULT 3

AAA97382

ID AAA97382 standard; cDNA; 3441 BP.

XX

AC AAA97382;

XX

DT 29-JAN-2001 (first entry)

XX

DE Pea light-repressible GTP-binding protein pra2 cDNA.

XX

KW GTP-binding protein pra2; pea; light-repressible promoter;

KW photoinhibitory; expression cassette; transgenic plant;

KW deterioration prevention; storage; ss.

XX

OS Pisum sativum.

XX

PN WO200055313-A1.

XX

PD 21-SEP-2000.

XX

PF 03-MAR-2000; 2000WO-JP01269.

XX

PR 12-MAR-1999; 99JP-0066551.

XX

PA (SUNR) SUNTORY LTD.

XX

PI Sasaki Y, Nagano Y, Inaba T;

XX

PS WPI; 2000-587526/55.

XX

DR P-PSDB; AAB23171.

XX

PT New DNA fragment or promoter for expressing a target gene, specifically

PT under photoinhibitory conditions, and for transforming a plant cell or

PT plant to improve quality and prevent deterioration during storage -

XX

PS Example 1; Fig 1; 49pp; Japanese.

XX

CC The invention relates to a light-repressible promoter (AAA97385), or

CC active fragments thereof (AAA97383, AAA97384), from the gene encoding

CC the pea GTP-binding protein pra2. The invention also relates to an

CC expression cassette containing the pra2 promoter or its active

CC fragments for the expression of a gene under photoinhibitory or dark

CC conditions in a plant, and to transgenic plants, their descendants

CC and plant tissues comprising the expression cassette. The expression

CC cassette of the invention can be used to generate transgenic plants in

CC which deterioration during storage in the dark is prevented. This is

CC particularly useful for agricultural products. The present sequence

CC represents the pea pra2 gene.

XX

SQ Sequence 3441 BP; 1143 A; 606 C; 540 G; 1152 T; 0 other;

```

Query Match      100.0%; Score 93; DB 21; Length 3441;
Best Local Similarity 100.0%; Pred. NO. 2.8e-14;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aaagtaacacatatatttgataaaatttattactaaactattttctagtagtctgttaac 60
    |||||
Db 1396 aaagtaacacatatatttgataaaatttattactaaactattttctagtagtctgttaac 1455
    |||||

QY 61 atgtctgaggattttacagtaataaagaacga 93
    |||||
Db 1456 atgtctgaggattttacagtaataaagaacga 1488

RESULT 4
AAF58252/c
ID AAF58252 standard; DNA; 936 BP.
XX
AC AAF58252;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1835.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
XX
PS 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match      36.1%; Score 33.6; DB 22; Length 936;
Best Local Similarity 3.3%; Pred. NO. 6.1;
Matches 3; Conservative 69; Mismatches 18; Indels 0; Gaps 0;

QY 1 aaagtaacacatatatttgataaaatttattactaaactattttctagtagtctgttaac 60
    :::::
Db 461 WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW 402

QY 61 atgtctgaggattttacagtaataaagaacga 90
    :::::
Db 401 WWWWWW TAAGC WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW 372

RESULT 5
AAF58252/c
ID AAF58252 standard; DNA; 936 BP.
XX
AC AAF58252;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1835.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX

```

```

AAF58254/c
ID AAF58254 standard; DNA; 936 BP.
XX
AC AAF58254;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1875.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
XX
PS 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

Query Match      36.1%; Score 33.6; DB 22; Length 936;
Best Local Similarity 3.3%; Pred. NO. 6.1;
Matches 3; Conservative 69; Mismatches 18; Indels 0; Gaps 0;

QY 1 aaagtaacacatatatttgataaaatttattactaaactattttctagtagtctgttaac 60
    :::::
Db 461 WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW 402

QY 61 atgtctgaggattttacagtaataaagaacga 90
    :::::
Db 401 WWWWWW TAAGC WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW 372

RESULT 6
AAF58257/c
ID AAF58257 standard; DNA; 936 BP.
XX
AC AAF58257;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1954.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX

```



```

XX OS Synthetic.
XX OS WO200107665-A2.
XX PN XX
XX PD 01-FEB-2001.
XX PF 26-JUL-2000; 2000WO-US20476.
XX PR 26-JUL-1999; 99US-0145695.
XX PR 17-MAR-2000; 2000US-0190259.
XX PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX PI Umek RM;
XX DR WPI; 2001-159728/16.
XX PT Nucleic acids containing electron-transfer group, useful as labels in hybridization assays, e.g. for genotyping, allowing repeat analyses on a single surface -
XX PS Example 4; Page 120; 159pp; English.
XX CC The present invention relates to a composition comprising two nucleic acids each containing an electron-transfer group (ETM) having different redox potentials. The invention is used for electronic detection of nucleic acids, especially of substitutions (mismatches) and single-nucleotide polymorphisms, e.g. for genotyping, monitoring gene expression.
XX SQ Sequence 244 BP; 19 A; 9 C; 12 G; 10 T; 194 other;

Query Match 35.3%; Score 32.8; DB 22; Length 244;
Best Local Similarity 1.1%; Pred. NO. 9.3;
Matches 1; Conservative 71; Mismatches 18; Indels 0; Gaps

Qy 1 aaagtaacacatatatttgataatttactaaaactattttctagtactgtgtaatc 60
Db 191 WWWWWWXXXXXXXXXXXXXWWWWWXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 61 atgtcgaggattttcacagtaataagaagaa 90
Db 131 WWWWWWXXXXXXXXXXXXXWWWWWXXXXXXXXXXXXXWWWWWXXXXXXXXXXXXX
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
AAF58252
ID AAF58252 standard; DNA; 936 BP.
XX AC AAF58252;
XX DT 24-APR-2001 (first entry)
XX DE Oligonucleotide D1835.
XX KW Electron-transfer group; ETM; mismatch; genotyping;
KW KW gene expression; ss.
XX OS Synthetic.
XX OS 
XX PN WO200107665-A2.
XX PD 01-FEB-2001.
XX PF 26-JUL-2000; 2000WO-US20476.
XX PR 26-JUL-1999; 99US-0145695.
XX PR 17-MAR-2000; 2000US-0190259.
XX PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX OS 

```


Qy 61 atgtctgaggattttacagtaataaagaaa 90
:: : : : : : : : : : : : : : : : : :
Db 82 wwwwwwwwwwwwwwwwwwwwwwwwwww 111

Search completed: April 3, 2002, 05:18:34
Job time: 2904 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2002, 05:18:34 ; Search time 189.71 Seconds
(without alignments)
10506.991 Million cell updates/sec

Title: US-09-700-187-3

Perfect score: 2325

Sequence: 1 aagctttaaggcaaggaa.....ttttgatcttgacaagaaa 2325

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 42862619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_1101.*

1: /SID52/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2325	100.0	2325	21 AAA97385	Pea pra2 gene ligh
2	2325	100.0	3441	21 AAA97382	Pea light-repressi
C 3	198.2	8.5	936	22 AAF58252	Oligonucleotide D1
C 4	198.2	8.5	936	22 AAF58254	Oligonucleotide D1
C 5	198.2	8.5	936	22 AAF58257	Oligonucleotide D1
C 6	198.2	8.5	936	22 AAF58259	Oligonucleotide D2
C 7	198.2	8.5	936	22 AAF58262	Oligonucleotide D1
C 8	198.2	8.5	936	22 AAF58255	Oligonucleotide D1
9	197.6	8.5	936	22 AAF58252	Oligonucleotide D1
10	197.6	8.5	936	22 AAF58254	Oligonucleotide D1
11	197.6	8.5	936	22 AAF58257	Oligonucleotide D1

12	197.6	8.5	936	22 AAF58259	Oligonucleotide D2
13	197.6	8.5	936	22 AAF58262	Oligonucleotide D1
14	197.6	8.5	938	22 AAF58255	Oligonucleotide D1
15	93	4.0	93	21 AAA97384	Pea pra2 gene ligh
16	75.4	3.2	244	22 AAF58238	Oligonucleotide D1
C 17	71.8	3.1	244	22 AAF58238	Oligonucleotide D1
18	53.2	2.3	5409	21 AAA70151	Plasmodium falcipa
19	52.4	2.3	513445	22 AA161373	Soybean 318013 reg
20	51.4	2.2	1664976	19 AAV21209	Methanococcus jann
21	50.4	2.2	19124	18 AAT72882	Plasmodium var-7 g
22	50.4	2.2	19124	21 AAZ98287	Plasmodium var-7 p
23	50.2	2.2	1627	14 AAQ47761	Antisense RNA comp
C 24	50.2	2.2	6152	18 AAT78867	P. falciparum live
25	49.4	2.1	1711	19 AAV33136	Plasmodium berghei
C 26	49.2	2.1	1230	10 AAN90223	Malaria-specific D
27	48.4	2.1	9789	17 AAT41852	cDNA encoding Plas
28	48.2	2.1	3975	9 AAN81157	Malaria-specific g
29	48.2	2.1	3975	13 AAQ22999	SERP gene. Plasm
30	48.2	2.1	6124	11 AAQ03568	Sequence encoding
31	47.2	2.0	1132	21 AAF12929	Aspergillus oryzae
C 32	46.8	2.0	2030	19 AAV29218	Nucleotide sequenc
C 33	46.4	2.0	3101	11 AAQ02047	Sequence encoding
C 34	46.4	2.0	10409	19 AAV42558	Mouse dectin-2 gen
35	46.2	2.0	441	21 AAC69763	Human breast tumo
36	46.2	2.0	53585	20 AAQ20251	Borrelia burgdorfe
C 37	46	2.0	488	21 AAC94232	Cat flea head and
38	46	2.0	161425	22 AAH02340	Human AKAP10 gene
39	46	2.0	162025	22 AAH02339	Human AKAP10 gene
40	45.8	2.0	376	22 AAH93354	Human chromosome 1
41	45.8	2.0	1431	21 AAZ37082	DNA sequence encod
42	45.8	2.0	1671	13 AAQ24134	50 kD subunit of S
43	45.8	2.0	3285	22 AAH54777	S. epidermidis gen
C 44	45.8	2.0	314	22 AAH54563	S. epidermidis gen
C 45	45.6	2.0	1078	21 AAQ02581	Human colon cancer

ALIGNMENTS

RESULT 1
AAA97385
ID AAA97385 standard; DNA; 2325 BP.
XX
AC AAA97385;
XX
DT 29-JAN-2001 (first entry)
XX
DE Pea pra2 gene light-repressible promoter.
XX
KW GTP-binding protein pra2; pea; light-repressible promoter;
KW photoinhibitory; expression cassette; transgenic plant;
KW deterioration prevention; storage; ds.
XX
OS Pisum sativum.
XX
PN WO200055313-A1.
XX
PD 21-SEP-2000.
XX
PF 03-MAR-2000; 2000WO-JP01269.
XX
PR 12-MAR-1999; 99JP-0066551.
XX
(SUNR) SUNTORY LTD.
XX
PI Sasaki Y, Nagano Y, Inaba T;
XX
DR WPI; 2000-587526/55.
XX
PT New DNA fragment or promoter for expressing a target gene, specifically
PT under photoinhibitory conditions, and for transforming a plant cell or
PT plant to improve quality and prevent deterioration during storage -

PS Claim 3; Page 34-35; 49pp; Japanese.

XX The invention relates to a light-repressible promoter (AAA97385), or
CC active fragments thereof (AAA97383, AAA97384), from the gene encoding
CC the pea GTP-binding protein pra2. The invention also relates to an
CC expression cassette containing the pra2 promoter or its active
CC fragments for the expression of a gene under photoinhibitory or dark
CC conditions in a plant, and to transgenic plants, their descendants
CC and plant tissues comprising the expression cassette. The expression
CC cassette of the invention can be used to generate transgenic plants in
CC which deterioration during storage in the dark is prevented. This is
CC particularly useful for agricultural products. The present sequence
XX represents the pea pra2 promoter.

SQ Sequence 2325 BP; 791 A; 451 C; 317 G; 766 T; 0 other;

Query Match 100.0%; Score 2325; DB 21; Length 2325;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	aagctttaagcgcaagggaagacacaattccaaaataataaaaaactctcgaagaatga	60
DB	1	aagctttaagcgcaagggaagacacaattccaaaataataaaaaactctcgaagaatga	60
QY	61	ttttattcttctcataataacttttctctattccaaaaacacatcaaaagtattgtga	120
DB	61	ttttattcttctcataataacttttctctattccaaaaacacatcaaaagtattgtga	120
QY	121	ttcatattcttaattatcgtataataataattgtatttcaataatttctacaaattgtg	180
DB	121	ttcatattcttaattatcgtataataataattgtatttcaataatttctacaaattgtg	180
QY	181	ttatatgaataattttgtagttaaaagggaactaagaataacctccgcaacatcaaaagtca	240
DB	181	ttatatgaataattttgtagttaaaagggaactaagaataacctccgcaacatcaaaagtca	240
QY	241	gaacaccttctgaactcttcagttgaacagagaaggaagtggaacacacagaaactaaag	300
DB	241	gaacaccttctgaactcttcagttgaacagagaaggaagtggaacacacagaaactaaag	300
QY	301	ttccccacttaactcttggtttggttgaggactctcttcaaatatttactctaagga	360
DB	301	ttccccacttaactcttggtttggttgaggactctcttcaaatatttactctaagga	360
QY	361	aatacatagacactctagatggtttgcattagctcatatatttttaagtaataataccc	420
DB	361	aatacatagacactctagatggtttgcattagctcatatatttttaagtaataataccc	420
QY	421	acttcaagttttttgtttttgtttgtgtgcatgataagatggtatctttctcaa	480
DB	421	acttcaagttttttgtttttgtttgtgtgcatgataagatggtatctttctcaa	480
QY	481	ggcccttatgcaagacataaagatccatataactccaccaagattgtttacatctaacca	540
DB	481	ggcccttatgcaagacataaagatccatataactccaccaagattgtttacatctaacca	540
QY	541	agttaatgaattttaaattcttcgaacaaatttttctaccacaaaggaagtattatgac	600
DB	541	agttaatgaattttaaattcttcgaacaaatttttctaccacaaaggaagtattatgac	600
QY	601	attttctaatgtatttttatagaattgtatacatgttttctgtttatacaagattagaatt	660
DB	601	attttctaatgtatttttatagaattgtatacatgttttctgtttatacaagattagaatt	660
QY	661	tggatttctcaatcaactctacacttggtagaataatttcagcctcaactcagtaaat	720
DB	661	tggatttctcaatcaactctacacttggtagaataatttcagcctcaactcagtaaat	720
QY	721	caggttctctcttcaactcatcacttgggtgagtagaattatggagcgtcaacttagc	780
DB	721	caggttctctcttcaactcatcacttgggtgagtagaattatggagcgtcaacttagc	780

QY	781	aatatgaatccctctctccaaagatcctacacttatctgtagtgagaaattttgtctcgcacct	840
DB	781	aatatgaatccctctccaaagatcctacacttatctgtagtgagaaattttgtctcgcacct	840
QY	841	caacaagatagatttgatgggtcatcacgaggggaaacattccacattgggtcacaagattc	900
DB	841	caacaagatagatttgatgggtcatcacgaggggaaacattccacattgggtcacaagattc	900
QY	901	accacaacaagtgtgagagacatcacatatcaaccacaaaccttaaggtgtatgtgtatg	960
DB	901	accacaacaagtgtgagagacatcacatatcaaccacaaaccttaaggtgtatgtgtatg	960
QY	961	agttcttacttataaaagtgtcaacctccacttttcttaagcaattgtgtacttagaac	1020
DB	961	agttcttacttataaaagtgtcaacctccacttttcttaagcaattgtgtacttagaac	1020
QY	1021	tcacacttatttctcaacataaactcacacttgttttatacaacaattctccccacaagtgtg	1080
DB	1021	tcacacttatttctcaacataaactcacacttgttttatacaacaattctccccacaagtgtg	1080
QY	1081	agttcattcgtctatgtccctcccaagtggaaotctcttccatcccgcatgcttataccgttg	1140
DB	1081	agttcattcgtctatgtccctcccaagtggaaotctcttccatcccgcatgcttataccgttg	1140
QY	1141	ttgacatacatcttactctgctatgggacgttcaattgggacacgctgctgaccacatg	1200
DB	1141	ttgacatacatcttactctgctatgggacgttcaattgggacacgctgctgaccacatg	1200
QY	1201	tcagaagacgttttgacacaaaggagtcgctccctactcgaaccagactctgtaccatt	1260
DB	1201	tcagaagacgttttgacacaaaggagtcgctccctactcgaaccagactctgtaccatt	1260
QY	1261	aatagatcaactttgaaatgggatatcattcatctataatacaacatttactgtaaaagataaa	1320
DB	1261	aatagatcaactttgaaatgggatatcattcatctataatacaacatttactgtaaaagataaa	1320
QY	1321	aattccccaaaaaaatgagagagacactacatctcttatttataataataaatgttaa	1380
DB	1321	aattccccaaaaaaatgagagagacactacatctcttatttataataataaatgttaa	1380
QY	1381	agaaaaatatagtaataaaagttaacacataattttgataaaatttacttaaaactatttcc	1440
DB	1381	agaaaaatatagtaataaaagttaacacataattttgataaaatttacttaaaactatttcc	1440
QY	1441	tagtacttgttaataatctgtctgaggattttacagtaataaagaacagaggttagcccaaac	1500
DB	1441	tagtacttgttaataatctgtctgaggattttacagtaataaagaacagaggttagcccaaac	1500
QY	1501	aaaagtgataattgtgaggggtgtgactcttctgctgggtgcaaaaaaatgaaaccccaactt	1560
DB	1501	aaaagtgataattgtgaggggtgtgactcttctgctgggtgcaaaaaaatgaaaccccaactt	1560
QY	1561	gtgatatgtgtcgactgctccgtcgtacattgaaataatgaattgtcttttataacg	1620
DB	1561	gtgatatgtgtcgactgctccgtcgtacattgaaataatgaattgtcttttataacg	1620
QY	1621	tttgtctatgcgtattaccatcatggtctactagaatgggacaaatgaatttataatat	1680
DB	1621	tttgtctatgcgtattaccatcatggtctactagaatgggacaaatgaatttataatat	1680
QY	1681	ctgtcatgtgtgggtggaattcaaatgtatcgtaaatgttctctccatgaattc	1740
DB	1681	ctgtcatgtgtgggtggaattcaaatgtatcgtaaatgttctctccatgaattc	1740
QY	1741	acacaattatcatcactggtcaactcactgggtcaattgttttctctccatgaattc	1800
DB	1741	acacaattatcatcactggtcaactcactgggtcaattgttttctctccatgaattc	1800
QY	1801	acattgtctaaagaaaaattaccacactttaaattgtttatcccttgcacacattttcacatcaa	1860
DB	1801	acattgtctaaagaaaaattaccacactttaaattgtttatcccttgcacacattttcacatcaa	1860
QY	1861	tttattaaaaacatttttaccattgggaaaaacacatacatattcaatcaattattttgtcatt	1920

Best Local Similarity 0.6%; Pred. No. 5.7e-29;
Matches 5; Conservative 552; Mismatches 230; Indels 0; Gaps 0;

```

Qy 1361 attattataataaagttaaagaaaaatatagttataaaagtaacacatatittgataaat 1420
Db 787 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 728

Qy 1421 ttattactaaacactattttctagttactttgttaactatgtctgagagattttacagtaataa 1480
Db 727 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 668

Qy 1481 agaaacagggtagcccaacaaacaaagtgataattgtgggggtgtgatcttctggtgtgca 1540
Db 667 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 608

Qy 1541 aaaaatgaaccaccaacttgtatattgtctgaactgtcgcgtcgctacattgaaatta 1600
Db 607 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 548

Qy 1601 atgaatgtcttttataacggtttgtctatgcgctattaccocatatggctcactagaatggg 1660
Db 547 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 488

Qy 1661 acaatgaatttaatatatctgtctgtggtgggtgcgaatttaattgtatcgtgtaaa 1720
Db 487 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 428

Qy 1721 tggtagacatactgctacacaatttatcatctactgctgctcaactgctgctcaatgtg 1780
Db 427 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 368

Qy 1781 ttctctctccatgaattccattgtctaaagaaaattaccacctttaaattgtttatccc 1840
Db 367 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 308

Qy 1841 ttgcacacatttcacatcaattttattaaaaacattttaccatttggaacacacatcatatt 1900
Db 307 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 248

Qy 1901 caatcaattattttgcatcttttcaaaaactaaacacaaacttagaataattttgtaatt 1960
Db 247 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 188

Qy 1961 tatagcaaatatttcaaaaatatcctagttctcaaccactcaataattcacaatttccaa 2020
Db 187 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 128

Qy 2021 atcccttgcaaaacatcacacacacctctagaaaactttgatttaataatctataaaaagcaata 2080
Db 127 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 68

Qy 2081 atatgatatctcaacaataatcacacatatatgttatgatataatatgctgcagcaatacac 2140
Db 67 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 8

Qy 2141 ttaattt 2147
Db 7 WWWWWWW 1

RESULT 4
AAF58254/c
ID AAF58254 standard; DNA; 936 BP.
XX
AC AAF58254;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1875.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
XX
KW gene expression; ss.
XX

```

[illegible]


```
XX PF 26-JUL-2000; 2000WO-US20476.
XX PR 26-JUL-1999; 99US-0145695.
XX PR 17-MAR-2000; 2000US-0190259.
XX PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX PI Umek RM;
XX DR WPI; 2001-159728/16.
XX PT Nucleic acids containing electron-transfer group, useful as labels in
XX PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX PT a single surface
XX PS Example 6; Page 128; 159pp; English.
XX CC The present invention relates to a composition comprising two nucleic
XX CC acids each containing an electron-transfer group (ETW) having
XX CC different redox potentials. The invention is used for electronic
XX CC detection of nucleic acids, especially of substitutions (mismatches)
XX CC and single-nucleotide polymorphisms, e.g. for genotyping,
XX CC monitoring gene expression.
XX SQ Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

Query Match 8.5%; Score 198.2; DB 22; Length 936;
Best Local Similarity 0.6%; Pred. No. 5.7e-29;
Matches 5; Conservative 552; Mismatches 230; Indels 0; Gaps 0;

Qy 1361 attatattaataaagttaaagaaataatagatatataaaagtaaacacatatattgataaat 1420
Db 787 WWWWWW... 728

Qy 1421 ttattactaaactatttcttagtactgttaatactgctgaggtatttacagtaataa 1480
Db 727 WWWWWW... 668

Qy 1481 agaaacgaggtagcccaacaaagtgataattgtggagggtgtagcttctgcggtgca 1540
Db 667 WWWWWW... 608

Qy 1541 aaaaatgaaccccaactgtgtattgtgcactgctgcgtcgctacattgaaatta 1600
Db 607 WWWWWW... 548

Qy 1601 atgaatgttctttataacgtttgtctatgcgctattaccocatatggctacataagtg 1660
Db 547 WWWWWW... 488

Qy 1661 acaatgaatttaataatatactgtcatgtggtgggtggtggttaatttatgtatgtaaa 1720
Db 487 WWWWWW... 428

Qy 1721 tggtaggacatactatgctacacaaattatcatcatcactggtcactggtcactgtg 1780
Db 427 WWWWWW... 368

Qy 1781 tttctctcccatgaattcatcattgttaaagaaattaccaccttaaaattgtttatccc 1840
Db 367 WWWWWW... 308

Qy 1841 ttgcacacatttcacatacaatttattataaaacattttacattggaacacacacataatt 1900
Db 307 WWWWWW... 248

Qy 1901 caatacaatttttgcatttttcaaaaactaaacacaaacaaacttagaatttttgtaatt 1960
Db 247 WWWWWW... 188

Qy 1961 tatagcaaattttcaaaaatatccttagtcttcaaccactcaataaattcacaatttccaa 2020
```

```
Db 187 WWWWWW... 128

Qy 2021 atcccttgcaaaaacatcacaaacctctagaaactttgataataatcttaataaaagcaata 2080
Db 127 WWWWWW... 68

Qy 2081 atatgatatacacaataatcacacatatgttatgataataatgatgcagcaatacac 2140
Db 67 WWWWWW... 8

Qy 2141 ttaattt 2147
Db 7 WWWWWW 1

RESULT 7
AAF38262/C
ID AAF58262 standard; DNA; 936 BP.
XX AC AAF58262;
XX DT 24-APR-2001 (first entry)
XX DE Oligonucleotide D2007.
XX KW Electron-transfer group; ETM; mismatch; genotyping;
XX KW gene expression; ss.
XX OS Synthetic.
XX PN WO200107665-A2.
XX PD 01-FEB-2001.
XX PF 26-JUL-2000; 2000WO-US20476.
XX PR 26-JUL-1999; 99US-0145695.
XX PR 17-MAR-2000; 2000US-0190259.
XX PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX PI Umek RM;
XX DR WPI; 2001-159728/16.
XX PT Nucleic acids containing electron-transfer group, useful as labels in
XX PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX PT a single surface
XX PS Example 6; Page 128; 159pp; English.
XX CC The present invention relates to a composition comprising two nucleic
XX CC acids each containing an electron-transfer group (ETW) having
XX CC different redox potentials. The invention is used for electronic
XX CC detection of nucleic acids, especially of substitutions (mismatches)
XX CC and single-nucleotide polymorphisms, e.g. for genotyping,
XX CC monitoring gene expression.
XX SQ Sequence 936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;

Query Match 8.5%; Score 198.2; DB 22; Length 936;
Best Local Similarity 0.6%; Pred. No. 5.7e-29;
Matches 5; Conservative 552; Mismatches 230; Indels 0; Gaps 0;

Qy 1361 attatattaataaagttaaagaaataatagatatataaaagtaaacacatatattgataaat 1420
Db 787 WWWWWW... 728

Qy 1421 ttattactaaactatttcttagtactgttaatactgctgaggtatttacagtaataa 1480
Db 727 WWWWWW... 668
```


[illegible]

RESULT 11
AAF58257

[illegible]

[illegible]

RESULT 15

ALSO IN
AAA97384
ID AAA97384 standard; DNA; 93 BP.

AA
AC
AAA97384:

DT 29-JAN-2001 (first entry)

DE Pea pra2 gene light-repressible promoter cis element.

KW GTP-binding protein pra2; pea; light-repressible promoter;
 KW photoinhibitory; expression cassette; transgenic plant;
 KW deterioration prevention; storage; cis element; ds.

OS Pisum sativum.

XX PN WO200055313-A1.

21-SEP-2000.

03-MAR-2000: 2000WO-JP01269.

XX 12-MAR-1999: 99JP-0066551.

PA (SUNR) SUNTORY LTD.

PI Sasaki Y, Nagano Y, Inaba T:

XX
I

DR XX PT PT PT XX PS XX CC CC CC CC CC CC CC CC CC CC CC XX SQ

WPI; 2000-587526/55.

New DNA fragment or promoter for expressing a target gene, specifically under photoinhibitory conditions, and for transforming a plant cell or plant to improve quality and prevent deterioration during storage -

Claim 2: Page 33; 49pp; Japanese.

The invention relates to a light-repressible promoter (AAA97385), or active fragments thereof (AAA97383, AAA97384), from the gene encoding the pea GRP-binding protein pra2. The invention also relates to an expression cassette containing the pra2 promoter or its active fragments for the expression of a gene under photoinhibitory or dark conditions in a plant, and to transgenic plants, their descendants and plant tissues comprising the expression cassette. The expression cassette of the invention can be used to generate transgenic plants in which deterioration during storage in the dark is prevented. This is particularly useful for agricultural products. The present sequence represents the pea pra2 promoter *cis* element.

Sequence 93 BP; 38 A; 10 C; 11 G; 34 T; 0 other;

Query Match 4.08; Score 93; DB 21; Length 93;

Best Local Similarity 100.0%; Pred. No. 3.7e-09;

Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1396 aaagtaacacataattttgataaattattactaaaactattttctagtagtacttgtaatc 1455

Db 1 aaagtaacacacataatttttgataaattattactaaaacatttttcagtaactattgaatc 600

Qy 1456 atgtctgaggattttacagtaataaaqaacqa 1488

[illegible]

Search completed: April 3, 2002, 05:19:04
Job time: 2934 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2002, 05:14:55 ; Search time 81.95 Seconds
(without alignments)
6425.398 Million cell updates/sec

Title: US-09-700-187-3
Perfect score: 2325
Sequence: 1 aagctttaaggaaggaa.....ttttgatcttgacaagaaa 2325

Scoring table: IDENTITY_NUC
Gapex 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50.4	2.2	19124	2	US-08-487-826B-13
c 2	50.2	2.2	6152	2	US-08-973-462-1
c 3	46.8	2.0	2030	2	US-08-705-937-7
c 4	46.4	2.0	10409	3	US-08-772-440-33
5	45.6	2.0	2861	1	US-08-299-953-1
6	45.6	2.0	2861	1	US-08-459-415-1
7	45.6	2.0	2861	5	PCT-US95-11231-1
8	45.6	2.0	3881	1	US-08-299-953-2
9	45.6	2.0	3881	1	US-08-459-415-2
10	45.6	2.0	3881	5	PCT-US95-11231-2
c 11	44.6	1.9	731	1	US-08-451-405A-2
c 12	44.2	1.9	1183	2	US-08-731-722-8
c 13	44	1.9	854	4	US-08-998-416-534
c 14	44	1.9	19124	2	US-08-487-826B-13
c 15	43.8	1.9	5852	1	US-07-867-106-2
c 16	43.6	1.9	1422	1	US-08-319-704-5
c 17	43.2	1.9	1850	3	US-08-617-860B-32
c 18	43.2	1.9	4098	2	US-08-605-106-4
c 19	43	1.8	665	2	US-08-883-795A-36
c 20	43	1.8	2781	3	US-08-749-522-4
c 21	43	1.8	7218	1	US-08-232-463-14
c 22	42.8	1.8	615	4	US-08-998-416-186
c 23	42.8	1.8	827	4	US-08-998-416-535
c 24	42.6	1.8	662	4	US-08-998-416-185
c 25	42.6	1.8	663	4	US-08-998-416-191
c 26	42.6	1.8	665	4	US-08-998-416-937
c 27	42.6	1.8	701	4	US-08-998-416-701

c 28	42.6	1.8	711	4	US-08-998-416-786	Sequence 786, App
c 29	42.6	1.8	724	4	US-08-998-416-683	Sequence 683, App
c 30	42.6	1.8	732	4	US-08-998-416-1036	Sequence 1036, App
c 31	42.6	1.8	767	4	US-08-998-416-472	Sequence 472, App
c 32	42.6	1.8	828	4	US-08-998-416-538	Sequence 538, App
c 33	42.6	1.8	834	4	US-08-998-416-305	Sequence 305, App
c 34	42.6	1.8	860	4	US-08-998-416-287	Sequence 287, App
c 35	42.4	1.8	636	4	US-08-998-416-1137	Sequence 1137, App
c 36	42.4	1.8	658	4	US-08-998-416-595	Sequence 595, App
c 37	42.2	1.8	615	4	US-08-998-416-186	Sequence 186, App
c 38	42.2	1.8	834	4	US-08-998-416-305	Sequence 305, App
c 39	42	1.8	1431	4	US-09-316-083-2	Sequence 2, Appli
c 40	41.8	1.8	5394	3	US-08-688-376-1	Sequence 1, Appli
c 41	41.6	1.8	8920	2	US-08-446-855A-1	Sequence 1, Appli
c 42	41.6	1.8	8920	4	US-09-150-741-1	Sequence 1, Appli
c 43	41.4	1.8	2287	4	US-08-845-258-8	Sequence 8, Appli
c 44	41.4	1.8	2287	4	US-08-990-571-8	Sequence 8, Appli
c 45	41.4	1.8	2287	4	US-08-723-142A-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1
US-08-487-826B-13
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Welles, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelsen, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 19124 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-487-826B-13

Query Match

2.2%; Score 50.4; DB 2; Length 19124;

```
Best Local Similarity 48.0%; Pred. No. 0.038;
Matches 144; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

QY 1831 tttttcccttgccacatttcacatcaatttattataaaacattttacatttggaanaa 1890
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 18094 ttttacttcaatttcgataaattccgagcgaacaaaaaataataatctcattataaanaatt 18153
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1891 catacatatcaatcaattatttttgctatttcaaaacactaaacccaaacaaacttagaat 1950
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 18154 atttataatcaatattatattagtttccctatttaaaataaatttaataatataatataat 18213
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1951 atttgttaattagcacaaattttccaaataatccctagctcttcaacccactcaaatattca 2010
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 18214 atttctgttatttttataaataataactaattttctatttttttttttaactttattcccttt 18273
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2011 caatttccaaatcccttgcaaacatcacaaacctctagaaactttgattgaataatccta 2070
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 18274 ttaatttcttaattcttttatcaacaaacaaacataaagtaattttacataatcaacaaaa 18333
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2071 aaaaacaaataatgatattctaaacaaatcacccatataatgttatgataataatgatgc 2130
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 18334 aaaaaaaacaaaaaaataattttattataataataaataaataaataaagacataac 18393
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 2
US-08-973-462-1/c
; Sequence 1, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 6152
; TYPE: DNA
; ORGANISM: P. falciparum
US-08-973-462-1

Query Match 2.2%; Score 50.2; DB 4; Length 6152;
Best Local Similarity 45.8%; Pred. No. 0.031;
Matches 210; Conservative 0; Mismatches 248; Indels 1; Gaps 1;

QY 1809 aaagaaaattaccacttaaaatgtttatcccttgccacacatttcacatcaattttatttaa 1868
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6052 aaaaaaaacaaaaaaagcttttttgcavpaaacgcaaaataagctaaattttttatc 5993
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1869 aacattttaccattggaacacacacatattcaatcaattatttttgcatcttcaaaaa 1928
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5992 ttctaataatataatataatataatataatataatataatataatataatataatataat 5933
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1929 ctaaacaaaacaaacttagaattttgttaattatagcacaaattttcaaaaaatactctag 1988
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5932 acaaaataaataattatataatataatataatataatataatataatataatataatataat 5873
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1989 tttcaaacactcaataattccaaatttccaaatcccttgcaaaaacacacacacacacctctag 2048
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5872 atatttaattta-aaaaacataaaactctgtaaaaaaacattttataattaaaaataatttg 5814
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2049 aaactttgtaataatctcaataaagcaataatgatataatcaacaaatcacaccat 2108
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5813 aaaaataaattatataatgcgaataaataaataaattatataatataatataatataatataa 5754
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2109 atgttatgataataatgatgcagcaacacacttaattttgttgaaagcagacagac 2168
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Db 5753 aattttgttaaaagatacatatatatatatatatatatatatatatatatatatatatttt 5694
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2169 aactctattaaacacgggtaattcaacaaccggtgtgtgtgcagagttcattgtttcttccaac 2228
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5693 atatatattttatcatataatatacatattttttatatatttttttatcgatttaattttttgtcgc 5634
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2229 tctttcccttcccttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 2267
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5633 tccaatcatttcactctctctctctctctctctctctctctctctctctctctctctctct 5595
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 3
US-08-705-937-7/c
; Sequence 7, Application US/08705937
; Patent No. 5981841
; GENERAL INFORMATION:
; APPLICANT: Santino, Colleen G.
; APPLICANT: Conner, Timothy W.
; TITLE OF INVENTION: EARLY SEED 5' REGULATORY SEQUENCE
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carmen Rodriguez, Paralegal, Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/705,937
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Nicolas G. Barzoukas.
; REGISTRATION NUMBER: 38,823
; REFERENCE/DOCKET NUMBER: MOBT:018 (38-2(10694))A
; NAME: Janelle D. Waack.
; REGISTRATION NUMBER: 36,300
; REFERENCE/DOCKET NUMBER: MOBT:018 (38-2(10694))A
; NAME: Barbara S. Kitchell
; REGISTRATION NUMBER: 33,928
; REFERENCE/DOCKET NUMBER: MOBT:018 (38-2(10694))A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (713) 787-1400
; TELEFAX: (713) 789-2679
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2030 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-705-937-7

Query Match 2.0%; Score 46.8; DB 2; Length 2030;
Best Local Similarity 52.0%; Pred. No. 0.12;
Matches 105; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY 1282 atcattcactatcatcaaacatttcacgtaagataaaaaaacattcccccaacaaatgaga 1341
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 826 agcgaataataatagagataatagtaaaaaataataataataataataataataataataatga 767
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1342 gagacactacatctctctctctctctctctctctctctctctctctctctctctctctct 1401
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 766 aaaaataaataaataaataaataaataaataaataaataaataaataaataaataaataaata 707
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1402 aacacatatatttgataaaattttactaaactattttcttagtacttgtaataatcatgtct 1461
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
,
,
, NAME/KEY: modified_base
, LOCATION: 3405..6871
, OTHER INFORMATION: /mod_base= OTHER
, OTHER INFORMATION: /note= "S = C or G"
,
, FEATURE:
```

```

;
; OTHER INFORMATION: /mod_base= OTHER
; OTHER INFORMATION: /note= "W = A or T"
;
; FEATURE:
;
; NAME/KEY: modified base
; LOCATION: 3595..9999
;
; OTHER INFORMATION: /mod_base= OTHER
; OTHER INFORMATION: /note= "Y = C or T"
;
US-08-772-440-33

```

Query Match	2.0%	Score 46.4	DB 3	Length 10409
Best Local Similarity	48.7%	Pred. No. 0.24		
Matches 153: Conservative	1	Mismatches 157	Indels 33	Gaps 1

Qy 1885 aaacacatacatattccaatcatttttgcatttcacaaactaaccaacaact 1944
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| : ||| |
Db 4502 AACAAACAAACAACAACTAAGTCTATRGGAAATAGACAAAGAAAAAGARGCCCAAGTT 4443

[illegible][illegible]

Db	4382	GATGATTATCAAGGTACAAAAAACATCAGAAACACCAAAATGGACTGGATCAGAAAAGATA	4323
Qy	2062	taatcctaataaaagcaataatgatatactataatcctaataatcaccataatgatgtatgatata	2121

Db 4322 TTCCCAACACACACAATATCAAAATGCTAAACATACAGAAACAAGAAAGAAATGTGAAA 4263

	Qy	2122	2123	2124	2125	2126	2127	2128	2129	2130	2131	2132	2133	2134	2135	2136	2137	2138	2139	2140	2141	2142	2143	2144	2145	2146	2147	2148	2149	2150	2151	2152	2153	2154	2155	2156	2157	2158	2159	2160	2161	2162	2163	2164	2165	2166	2167	2168	2169	2170	2171	2172	2173	2174	2175	2176	2177	2178	2179	2180	2181	2182	2183	2184	2185	2186	2187	2188	2189	2190	2191	2192	2193	2194	2195	2196	2197	2198	2199	2200	2201	2202	2203	2204	2205	2206	2207	2208	2209	2210	2211	2212	2213	2214	2215	2216	2217	2218	2219	2220	2221	2222	2223	2224	2225	2226	2227	2228	2229	2230	2231	2232	2233	2234	2235	2236	2237	2238	2239	2240	2241	2242	2243	2244	2245	2246	2247	2248	2249	2250	2251	2252	2253	2254	2255	2256	2257	2258	2259	2260	2261	2262	2263	2264	2265	2266	2267	2268	2269	2270	2271	2272	2273	2274	2275	2276	2277	2278	2279	2280	2281	2282	2283	2284	2285	2286	2287	2288	2289	2290	2291	2292	2293	2294	2295	2296	2297	2298	2299	2300	2301	2302	2303	2304	2305	2306	2307	2308	2309	2310	2311	2312	2313	2314	2315	2316	2317	2318	2319	2320	2321	2322	2323	2324	2325	2326	2327	2328	2329	2330	2331	2332	2333	2334	2335	2336	2337	2338	2339	2340	2341	2342	2343	2344	2345	2346	2347	2348	2349	2350	2351	2352	2353	2354	2355	2356	2357	2358	2359	2360	2361	2362	2363	2364	2365	2366	2367	2368	2369	2370	2371	2372	2373	2374	2375	2376	2377	2378	2379	2380	2381	2382	2383	2384	2385	2386	2387	2388	2389	2390	2391	2392	2393	2394	2395	2396	2397	2398	2399	2400	2401	2402	2403	2404	2405	2406	2407	2408	2409	2410	2411	2412	2413	2414	2415	2416	2417	2418	2419	2420	2421	2422	2423	2424	2425	2426	2427	2428	2429	2430	2431	2432	2433	2434	2435	2436	2437	2438	2439	2440	2441	2442	2443	2444	2445	2446	2447	2448	2449	2450	2451	2452	2453	2454	2455	2456	2457	2458	2459	2460	2461	2462	2463	2464	2465	2466	2467	2468	2469	2470	2471	2472	2473	2474	2475	2476	2477	2478	2479	2480	2481	2482	2483	2484	2485	2486	2487	2488	2489	2490	2491	2492	2493	2494	2495	2496	2497	2498	2499	2500	2501	2502	2503	2504	2505	2506	2507	2508	2509	2510	2511	2512	2513	2514	2515	2516	2517	2518	2519	2520	2521	2522	2523	2524	2525	2526	2527	2528
--	----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

Oy	2182	ccggttaattcaaca	2195
Db	4202	CCTCTCATTGAAGA	4189

RESULT 5

US-08-299-953-1
: Sequence 1, Application US/08299953
: Patent No. 5646333
: GENERAL INFORMATION.

GENERAL INFORMATION:
APPLICANT: Dobres,
TITLE OF INVENTION: A plant promoter useful for directing the expression of foreign proteins to the plant epidermis
TITLE OF INVENTION: Expression of foreign proteins to the plant epidermis

TITLE OF INVENTION:	expression of foreign proteins to the plant epidermis
NUMBER OF SEQUENCES:	4
CORRESPONDENCE ADDRESS:	Kurtz Mackiewicz & No. 5646333tris
ADDRESSEE:	WOODCOCK

STREET: One Liberty Place 46th. Floor
CITY: Philadelphia
STATE: PA

```
: ZIP: 19103  
:  
: COMPUTER READABLE FORM:  
:  
: MEDIUM TYPE: Floppy disk
```

```

: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:

```

```

:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/299,953
: FILING DATE: Herewith
: CLASSIFICATION: 435
:

```

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Beardell, Lori Y.


```
Db 1037 CATATGATAATACCCCTATCATTTTTTTTATATATGTTGTTTGAATAAATAATATATCAT 1096
Qy 508 tatactcaccagaagctgtttacatctcaaccaagttcaatgaattttaaattcttcgaaac 567
Db 1097 AATATAATCATTTTACAATGTTAAAGAGATTAATATTAGTTTTTTTATATATACTCTAG 1156
Qy 568 aattattctcaccagaagtttatcatgcacattttcttaagtatttttatataagaat 627
Db 1157 AATTTTATTAATATCTTTTTTTTAAATATATATAAATTTATCTCTCCATTTTATTAATAT 1216
Qy 628 tgatacatgtttctgatatcaagaattgaatttggatttctcaccacactccta 683
Db 1217 ATAAAGATATTTTTTTTGATATAATTTTTTAAATACGTGTAATAAATCTAAATATTA 1272

RESULT 7
PCT-US95-11231-1
; Sequence 1, Application PC/TUS9511231
; GENERAL INFORMATION:
; APPLICANT: Dobres, Michael S. and Mandaci, Sevnur
; TITLE OF INVENTION: A Plant Promoter Useful for Directing the Expression
; of Foreign Proteins to the Plant Epidermis
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & Norris
; STREET: One Liberty Place 46th. Floor
; CITY: Philadelphia
; STATE: PA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/11231
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/299,953
; FILING DATE: September 2, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Beardell, Lori Y.
; REGISTRATION NUMBER: 34,293
; REFERENCE/DOCKET NUMBER: NOVA-0016
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-564-8960
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2861 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
PCT-US95-11231-1

Query Match 2.0%; Score 45.6; DB 5; Length 2861;
Best Local Similarity 44.1%; Pred. No. 0.25;
Matches 289; Conservative 0; Mismatches 359; Indels 8; Gaps 2;

Qy 34 aaaaataaaaactcctaagaagatgtttttattcttattcttctcaataaacttttcta 93
Db 619 AATTTGTTTAGATATATTAATAAATAAATATGCGCAATATTTTTTAAAGATCAGACTAA 678
Qy 94 ttccaaaacacatcaagttgtgtcatcttcttaattctctgataataataatt 153
Db 679 GAGAAAAAACAATATGATAT--TTATTAACATTTTATGCTTTAAATAAATGATT 736
Qy 154 gtatttcaatattctcacaattgtgttatgaaatattttaggttaaaaggacta 213
```

```
Db 737 AASTTTTTATAAAAAATTAATCTAAAAATGATAAAATTAATATTTTAATATTGATTT 796
Qy 214 agaataacctccgcaacatcaaaagtcagaacctctgttaactcttcagttgaaacgaga 273
Db 797 AATACTGCATATCACCATTATTACTAACATGTCTATTAATAAATAGATTTTAAAAAGTAA 856
Qy 274 aggaagtggcacaacacagaacaaactaaagttcccccacttaactctctgtgttgggtgagga 333
Db 857 TCTAAAGAATATAAATAATTAATATTTTAATATCGACTCGAATTAGTAGTTAGTTGGACAC 916
Qy 334 ctctcttacaatttatactctaaaggaataacattagacactctagatgggttgcattag 393
Db 917 ATTGAATACTAGCTTTTATGACACTCCCAATTAGTTTGACACATTTGAATACTAGCTTTTCAT 976
Qy 394 ctcatatatttttaagtaataataccccacttcaagttttttttttttttttttttgtgtgtgcag 453
Db 977 TTCCAACCAAAATTTGTATCATTTTCCAAACCAAACTTGTCTTTTATTAATAATTTCTGA 1036
Qy 454 tagatgataaga-----tggatcatttctcaagggcccttatgcaagacataagaatcca 507
Db 1037 CATATGATAATACCTCCCTATCATTTTTTATTTATGTTGTTTGTGAAAAAATAATATATCAT 1096
Qy 508 tatactcaccagaagctgtttacatctcaaccaagtttaataagattttaaattcttcgaaac 567
Db 1097 AATATAAATCATTTTACAATGTTAAAGAGATTAATATTAGTTTTTTTATTAATAAATCTAG 1156
Qy 568 aattattctcaccagaagtttatcatgcacattttcttaagtatttttatataagaat 627
Db 1157 AATTTTATTAATATCTTTTTTTTAAATATATAAATTTATCTCTCCATTTTATTAATAT 1216
Qy 628 tgatacatgtttctgatatcaagaattgaatttggatttctcaccacactccta 683
Db 1217 ATAAAGATATTTTTTTTGATATAATTTTTTAAATACGTGTAATAAATCTAAATAATTA 1272

RESULT 8
US-08-299-953-2
; Sequence 2, Application US/08299953
; Patent No. 5646333
; GENERAL INFORMATION:
; APPLICANT: Dobres, Michael S. and Mandaci, Sevnur
; TITLE OF INVENTION: A Plant Promoter Useful for Directing the
; Expression of Foreign Proteins to the Plant Epidermis
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5646333ris
; STREET: One Liberty Place 46th. Floor
; CITY: Philadelphia
; STATE: PA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/299,953
; FILING DATE: Herewith
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Beardell, Lori Y.
; REGISTRATION NUMBER: 34,293
; REFERENCE/DOCKET NUMBER: NOVA-0003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-564-8960
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3881 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
```

; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-299-953-2

Query Match 2.0%; Score 45.6; DB 1; Length 3881;
Best Local Similarity 44.1%; Pred. No. 0.27;
Matches 289; Conservative 0; Mismatches 359; Indels 8; Gaps 2;

```
Oy 34 aaaaataataaaacccctaaagaatgattttattcttcttcataaaataacttttcccta 93
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 619 AAATTTGGTTATAGATATTAAAAAATAATATGCGCAATATTTTTTAAAGATCAGACTAA 678

Oy 94 ttccaaaacacatcaaaagttatgtgattcatatttcaattatcttgataataataa 153
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 679 GAAGAAAAAACATTAATCTGATAT--TTATTAACATTTATTTGTTTAAATAAATGATT 736

Oy 154 gtatttcaataatttcatacaattgtgttatgataatatttttgagtaaaaggacta 213
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 737 AASTTTTTTATAAAAAAATAATCTAAAAATGATATAAATTAATTTTAAATATTGATT 796

Oy 214 agaaataaccccgcaacatacaagtcagaacccctctgttaacctcttcagttgaacgaga 273
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 797 AATCTGCGCATATCACCATTATTAATAATGCTGACTCGAATTAAGATTTAAAAAGTAA 856

Oy 274 aggaagtggacaacacagaaactaagttcccccacttaacctcttgggttgaggga 333
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 857 TCTAAAAAGAATATAAATAATATTTTAAATATCGACTCGAATTAAGTATTGAGTGGACAC 916

Oy 334 ctccctttacaattttactctaaagaaaatacatcttagacactctagatgggttgactag 393
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 917 APTGAATACTAGCTTTTATGGCACCTCCAATTAAGTTTGACACATTAAGTACTAGCTTTT 976

Oy 394 ctcatatttttaagtaataataaccacttaagtttttttttttttttttttttttttttt 453
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 977 TTCCAAACCAATTTGTATCATTTCCAAACCAACCTTGTCTTTTATAAATAAATTTCTGA 1036

Oy 454 tagatgataaga-----tgatcatcttctcaagcccttatgcaaaacacataagatcca 507
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1037 CATATGATAAATACTCCCTATCATTTTAAATAATGCTGTTTGAATAAATAATATATCAT 1096

Oy 508 tatactccaccaagatttgcatttacaataagtttgcacattttctaatgtattttatagaa 567
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1097 AATATAAATCATTTTACAAATGCTAAAAAGAAATTAATATTAGTTTATATAAATCTAG 1156

Oy 568 aattatttccctaccaagaggttttatgacattttctaatgtattttatagaa 627
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1157 AATTTTATTATATCTTTTTTTTAAATAATATAAATAATATCTCTCCCAATTTTATTAAATAT 1216

Oy 628 tgatacatgtttctgttatacaagattgaatttggatttctcatccaactccta 683
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1217 ATAAAGATATTTTGTATATAATTTTTTAAATACGTGTAATAAATCTAAATATTA 1272
```

RESULT 9

US-08-459-415-2
; Sequence 2, Application US/08459415
; Patent No. 5744334
; GENERAL INFORMATION:
; APPLICANT: Dobres, Michael S. and Mandaci, Sevnuur
; TITLE OF INVENTION: A Plant Promoter Useful for Directing the
; Expression of Foreign Proteins to the Plant Epidermis
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5744334ris
; STREET: One Liberty Place 46th. Floor
; CITY: Philadelphia
; STATE: PA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,415
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/299,953
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Beardell, Lori Y.
; REGISTRATION NUMBER: 34,293
; REFERENCE/DOCKET NUMBER: NOVA-0003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-564-8960
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3881 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-459-415-2

Query Match 2.0%; Score 45.6; DB 1; Length 3881;

Best Local Similarity 44.1%; Pred. No. 0.27;
Matches 289; Conservative 0; Mismatches 359; Indels 8; Gaps 2;

```
Oy 34 aaaaataataaaacccctaaagaatgattttattcttcttcataaaataacttttcccta 93
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 619 AAATTTGGTTATAGATATTAAAAAATAATATGCGCAATATTTTTTAAAGATCAGACTAA 678

Oy 94 ttccaaaacacatcaaaagttatgtgattcatatttcaattatcttgataataataa 153
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 679 GAAGAAAAAACATTAATCTGATAT--TTATTAACATTTATTTGTTTAAATAAATGATT 736

Oy 154 gtatttcaataatttcatacaattgtgttatgataatatttttgagtaaaaggacta 213
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 737 AASTTTTTTATAAAAAAATAATCTAAAAATGATATAAATTAATTTTAAATATTGATT 796

Oy 214 agaaataaccccgcaacatacaagtcagaacccctctgttaacctcttcagttgaacgaga 273
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 797 AATCTGCGCATATCACCATTATTAATAATGCTGACTCGAATTAAGATTTAAAAAGTAA 856

Oy 274 aggaagtggacaacacagaaactaagttcccccacttaacctcttgggttgaggga 333
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 857 TCTAAAAAGAATATAAATAATATTTTAAATATCGACTCGAATTAAGTATTGAGTGGACAC 916

Oy 334 ctccctttacaattttactctaaagaaaatacatcttagacactctagatgggttgactag 393
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 917 APTGAATACTAGCTTTTATGGCACCTCCAATTAAGTTTGACACATTAAGTACTAGCTTTT 976

Oy 394 ctcatatttttaagtaataataaccacttaagtttttttttttttttttttttttttttt 453
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 977 TTCCAAACCAATTTGTATCATTTCCAAACCAACCTTGTCTTTTATAAATAAATTTCTGA 1036

Oy 454 tagatgataaga-----tgatcatcttctcaagcccttatgcaaaacacataagatcca 507
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1037 CATATGATAAATACTCCCTATCATTTTAAATAATGCTGTTTGAATAAATAATATATCAT 1096

Oy 508 tatactccaccaagatttgcatttacaataagtttgcacattttctaatgtattttatagaa 567
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1097 AATATAAATCATTTTACAAATGCTAAAAAGAAATTAATATTAGTTTATATAAATCTAG 1156

Oy 568 aattatttccctaccaagaggttttatgacattttctaatgtattttatagaa 627
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1157 AATTTTATTATATCTTTTTTTTAAATAATATAAATAATATCTCTCCCAATTTTATTAAATAT 1216

Oy 628 tgatacatgtttctgttatacaagattgaatttggatttctcatccaactccta 683
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```


Query Match	1.98;	Score 44.2;	DB 2;	Length 1183;
Best Local Similarity	40.7%;	Pred. No. 0.4;		
Matches 155;	Conservative 0;	Mismatches 223;	Indels 3;	Gaps 17;
Qy 60	atttattcttattcctcctcaataaacttttctctattccaaaacacatcaaaagtattgtg	119		
Db 411	AATTAATTTATATCTTTTAAAGATATATAACTTTTAAAAAAAAATTTAAATGAAAAATA	470		
Qy 120	attcatattcttaattatctgtaataataatgttatattcaattatttcatacaaatgtt	179		
Db 471	CTTCTAATAAGAAATATATTTTATATAAATATTAATTATTTAAATTTACAAATCCATCA	530		
Qy 180	gttatatgaataattttgttaggttaaaagggactaagaataaacttcgcgaacatcaaatc	239		
Db 531	GAATATTTTAACTGTTATATAATATTAAAGAACTTTTACTTATATTTTCTAAATAAGACTGA	590		
Qy 240	agaaacctcttgttaactcttcagtttgaaacgagagaagggaagtggaacacacagaaactaa	299		

Query Match 1.98; Score 44; DB 4; Length 854;
Best Local Similarity 51.58; Pred. No. 0.4;
Matches 101; Conservative 0; Mismatches 95; Indels 0; Gaps 0;


```
? TELEFAX: 215-568-3439
? INFORMATION FOR SEQ ID NO: 2:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 5852 base pairs
? TYPE: NUCLEIC ACID
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? ANTI-SENSE: NO
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 2378..5038
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 2378..5038
? US-07-867-106-2

Query Match          1.9%; Score 43.8; DB 1; Length 5852;
Best Local Similarity 46.8%; Pred. No. 0.74;
Matches 171; Conservative 0; Mismatches 192; Indels 2; Gaps 1;

QY 1921 ttcaaaactaaacaaacaaacttagaataattttgtaattatagcacacaattttcaaaaa 1980
   ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 5312 TTTAAAAAAATTTTAAACACATGGAATATATAGATCGATAGATCACTAATTTTAA 5371

QY 1981 tatcctagcttcaaccactcaataattccaaatcccttgcaaaacacacaca 2040
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 5372 AATTAAATATATTAAATTTATAAAAATTCAGTTCATCAAGATATATAGATAATTATTTA 5431

QY 2041 accttagaaactttgatttaatactataaaagcaataatgatatactaaacaatat 2100
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 5432 ATTATTTGAATTTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 5491

QY 2101 caccatatatgtatgataataatgatgcagcaacacacctaatttggtaagcattaa 2160
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 5492 TTTAGATTTTAAATCTCGTCAATGATTTTAAATAAAAATCGATACATAA--TTTTAA 5549

QY 2161 agcgaacacactctataacaccggtaactcaacacccgttggtgcgagttcatgttt 2220
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 5550 AAAAAACCCCTTTACATTTTATTTTAAATCCAAATTTATACATTTTATTTTATTTT 5609

QY 2221 ctccaactctttccctttcttacttattatttctcctacttaocttttctacta 2280
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 5610 TTTTATTTTATTTTAAATTTAAATTTTATTTTATTTTATTTTATTTTATTTTAA 5669

QY 2281 atata 2285
   || | |
DB 5670 AATTA 5674
```

Search completed: April 3, 2002, 05:16:12
Job time: 3982 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2002, 04:39:16 ; Search time 2120.21 Seconds
(without alignments)
11783.708 Million cell updates/sec

Title: US-09-700-187-3

Perfect score: 2325

Sequence: 1 aagctttaaggcaaggga.....ttttgatctgtacaagaaa 2325

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*
1: em_estfun.*
2: em_esthum.*
3: em_estin.*
4: em_estom.*
5: em_estpl.*
6: em_estba.*
7: em_estro.*
8: em_estov.*
9: em_htc.*
10: gb_estl.*
11: gb_est2.*
12: gb_htc.*
13: gb_gss.*
14: em_gss_fun.*
15: em_gss_hum.*
16: em_gss_inv.*
17: em_gss_pln.*
18: em_gss_pro.*
19: em_gss_rod.*
20: em_gss_vrt.*
21: em_gss_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	91.6	3.9	1101	13	CNS0039G
2	91	3.9	1101	13	CNS00L2
3	79.8	3.4	1101	13	CNS00EVL
4	75.6	3.3	1101	13	CNS0021J
5	74	3.2	1101	13	CNS0182P
6	72.4	3.1	1101	13	CNS00EVL
7	71.4	3.1	587	10	AW774987
8	70.8	3.0	1101	13	CNS006TE
9	70.6	3.0	1125	13	AL547503
10	70	3.0	928	13	CNS00DKY
11	70	3.0	1101	13	CNS00E07
12	70	3.0	1101	13	CNS00EPO
					AL063921 Drosophil
					AL078714 Drosophil
					AL069706 Drosophil
					AL061936 Drosophil
					AL108811 Drosophil
					AL069706 Drosophil
					AW774987 EST334138
					AL065901 Drosophil
					AL547503 AL547503
					AL071865 Drosophil
					AL069440 Drosophil
					AL069493 Drosophil

13	70	3.0	1204	13	CNS016E2
14	69.4	3.0	1061	13	CNS015LM
15	69	3.0	759	13	CNS06QXV
16	68.2	2.9	1101	13	CNS00CYH
17	67.4	2.9	928	13	CNS00DKY
18	67.2	2.9	1101	13	CNS000D1
19	67	2.9	1101	13	CNS0039G
20	66.4	2.9	694	13	A0853360
21	66.2	2.8	963	13	CNS00A4L
22	65.8	2.8	1101	13	CNS002FG
23	65.8	2.8	1101	13	CNS016LI
24	64.6	2.8	734	13	CNS010MP
25	64.6	2.8	1131	13	CNS034FO
26	64.4	2.8	1101	13	CNS00399
27	64.2	2.8	1146	13	CNS021G2
28	64	2.8	1201	13	CNS0162X
29	63.8	2.7	1101	13	CNS00EJ4
30	63.8	2.7	1101	13	CNS0100X
31	63.6	2.7	1190	13	CNS020N7
32	63.4	2.7	693	10	AV682300
33	63.4	2.7	893	13	CNS013XE
34	63.4	2.7	1101	13	CNS00EPO
35	63.2	2.7	1101	13	CNS017KE
36	62.8	2.7	609	13	CNS025K2
37	62.8	2.7	987	13	CNS014PQ
38	62.6	2.7	942	13	CNS018GS
39	62.6	2.7	1201	13	CNS0165X
40	62.6	2.7	1225	13	CNS0161D
41	62.2	2.7	643	13	B11768
42	62	2.7	1201	13	CNS0167M
43	61.8	2.7	859	13	CNS004YV
44	61.8	2.7	889	13	CNS006W4
45	61.6	2.6	963	13	CNS0075X

ALIGNMENTS

RESULT 1

CNS0039G
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CNS0039G 1101 bp DNA GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
AL063921 GI:4941778
GSS.
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequençage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Oseogawa and Aaron Mammos in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.


```

RESULT      8
CNS006TE   1101 bp      DNA      GSS      03-JUN-1999
LOCUS       Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACR14L09 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL065901
VERSION     AL065901.1  GI:4944869
KEYWORDS    GSS.
SOURCE      fruit fly,
            Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 1101)
            Genoscope.
            Direct Submission
            Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage ;
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
            - Web : www.genoscope.cns.fr)
            Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mamoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library's
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             location/qualifiers
            source          1..1101
                        /organism="Drosophila melanogaster"
                        /db_xref="taxon:7227"
                        /clone_lib="RPCI-98"
                        /clone="BACR14L09"
                        /notes="end : TET3"
BASE COUNT      417 a   102 c   21 g   232 t   329 others
ORIGIN
Query Match      3.0%; Score 70.8; DB 13; Length 1101;
Best Local Similarity 30.1%; Pred. No. 0.078;
Matches 205; Conservative 128; Mismatches 341; Indels 7; Gaps 1;

QY 1291 actatcaaacatttcagtaaagataaaataattccaccagaaacaaagagagacacta 1350
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 400 ANNNNTAAAAAATAAAAAAATAAAAAAATAAAAAAATAATNAAATNNNTNCNTATNCCCTA 459
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1351 catctctctattattataataaaatgtaagaaaaatatagtaaaagtaaacacatat 1410
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 460 TATCTTCTTAATATNTNCCAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 519
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1411 ttgataaatttactaaaactatttcttagtactgttaactgtaactgtcgaggatttt 1470
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 520 AATATANTNNNTNTNTNTTNTTNTTNTTNTTNTTNTTTCGNTKGTNKTGAAATAA 579
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1471 acagtaaaagaacagaggtagcccaacaaaagtataattgtggagggtgtgacttt 1530
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 580 AAAANAGWNTNRTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 639
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1531 tgtcgggtgcaaaaaatgaaaccccaactgtgtatattgtgtcgactgctcgctgctac 1590
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 640 BMTAKGTSHGKWKVTMBKTBTBMDRAKWNVDNNSMTMNSTKWTYTTNNMYWTHHTATRT 699
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1591 attgaaattaatgaattcttttataacggtttgtcctatgcgcg-----tattaccoat 1643
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 700 KTTAYATHTAYKTKTYTCTTMTATBTWTYTTMTAYRMMAASATAATAAYBACAYMNA 759

```

BASE COUNT	262 a	70 c	84 g	321 t	191 others
ORIGIN					
Query Match	3.0%; Score 70; DB 13; Length 928;				
Best Local Similarity	33.5%; Pred. No. 0.11;				
Matches 120; Conservative	78; Mismatches 160; Indels 0; Gaps 0;				
Qy	1941	aacttagaatatttggtaattatagacaattttccaataatctcctagtcttcacacct	2000		
Dd	571	WAWTTTAAATWTATWTATWTAAAWTAWWATTATATWTATWAATAWAAATATWTATA	630		
Qy	2001	caataatcacaaatttccaaatcccttgcaaacatcacacotctagaaccttggatta	2060		
Dd	631	TAAATATWTWTWTATWTATATATAAATAAAAAAATAAAAAAATAAAAAATTTTAA	690		
Qy	2061	ataatctataaaaagcaataatatgatcatctaaacaatatcaccatatgttatgatat	2120		
Dd	691	WTAAAAAATAAAATTAATTTTTTTTTTTTTTTTATWTATWATAWAAATATWWWMAAT	750		
Qy	2121	aatatgatgcagcacacacttaatttggtaaagcattaaagcgagacactcattaac	2180		
Dd	751	WDGKNWNNAWWWWWWWWWWAAAATAAAAAAATAAAAAAATAAAAAAATAAAAA	810		
Qy	2181	accggtatccaacaacggttggctgagtcgatgtttctctccaactctttcccttt	2240		
Dd	811	WDDDDDDDDWKAACKKKKKKKKKKKKKTCTTKTTTGARWMTTTTTTTTTTTT	870		
Qy	2241	cctttactatttatttctctactactacotttctactactataataatactctctt	2298		
Dd	871	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	928		
RESULT 11					
CNS00E07/c					
LOCUS	CNS00E07	1101 bp	DNA	GSS	04 JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC: BACR29P01 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	AL069440				
KEYWORDS	AL069440.1	GI:4949583			
SOURCE	GSS.				
ORGANISM	Drosophila melanogaster fruit fly. Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
REFERENCE	1 (bases 1 to 1101) Genoscope.				
AUTHORS	Direct Submission				
TITLE	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)				
JOURNAL	- web : www.genoscope.cns.fr				
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACS. For further information please see http://www.fruitfly.org/TheBDGP melanogaster BAC library was prepared by Kazutoyo Osoeawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm .				
FEATURES	Location/Qualifiers				
source	1..1101 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone_lib="RPCI-98"				

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the *Drosophila*
melanogaster genome using these BACs. For further information
please see [http://www.fruitfly.org/The BDGP Drosophila](http://www.fruitfly.org/The%20BDGP%20Drosophila)
melanogaster BAC library was prepared by Kazutoyo Osoe and
Aaron Mammos in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of *Drosophila* DNA provided by the BDGP from the
isogenic strain y²; cn bw sp, the same strain used for the BDGP's
PI and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES	source	Location/Qualifiers
		1. .1101
		/organism="Drosophila melanogaster"
		/db_xref="taxon:7227"
		/clone_lib="pCI-98"
		/clone="BACR29p01"
		/note="end : T7"
BASE COUNT	289 a 155 c	307 t 232 others

Query Match	3.0%;	Score 70;	DB 13;	Length 1101;
Best Local Similarity	41.3%;	Pred. No. 0.1;		

[illegible]

Qy	1968	caatttcaaataatcctagttctcaaccactcaataattccaattccaaacacctt : : : :	2077
----	------	----------------------------------------------------------------------------------------------------------------------	------

Qy 2028 gaaaacatcacacacctctagaaactttgattaataataatctataaaaaagcaataatgat 2087

Qy 2088 atctaaacaatatcaccataatggttatgatataaatgatgcagcaacacttaattt 2147
| : | | | | : ||||| | | | | : | : | : | :

QY 2148 ggtaaagcattaaagcgagacaacacctattataacacccggtaatccaacacccgttggtc 2207

Qy 2208 gagttcagttttcttccaaactcttttcccttttcccttactttattttattctctactt 2267
| :: ||| : || | : | : : : || || ||| : :

Qy 2268 accttttctactaatata 2287
| ||| | : || |
| ||| | : || |

RESULT 13

LOCUS	DEFINITION	CNS016E2	1204 bp	DNA	GSS	26-JUL-1999
	Drosophila melanogaster genome survey sequence T7 end of BAC BACN15A12 of DrosBAC library from Drosophila melanogaster (fruit fly)					

ACCESSION	AL106628
VERSION	AL106628.1
KEYWORDS	GI:5622852 GSS.

ORGANISM Plasmid *Drosophila melanogaster*
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

[illegible]

Qy	1772	gtcaatgtgtttctctcccatgaattcacatgtcctaaagaaatttaccaccttaaa	1830
Db	908	WAAAACATTTTWTAAAWTAAATTTTAAWTTTACATTTTAAWTTAAWTTAAATTTAAATTTAA	849
Qy	1831	tgtttacccttgcacacattccacatcaattattataaacattttaccatgtggaaca	1890
Db	848	WAAAAARWATTAATTTTAAATTTTAAWTTAAWTTAAWTTAAWTTATTTTMMWAAATTTTAAWNA	789

D_b 788 TTTTCTTTTWTAAAGAWTTTWTATWATTAATTTTAAATWTTTAAATAAAAAAATTATTTTT 729

Db 728 TATTTWTATTAAGAAWTTTWTWTTTTAAATTTWYTTTAAATWTTTAAAWAWWTA 669

D_b 668 AWTAAATAAAATWTATTTAAATAATWAAATATAAAATTAAGDWTATATAWTTTTTTTTTWWC 609

Db 608 WTTWAATATATAAAATWAAAAATTAAATTTTATATWNNATAAAAAAMATWNNWTATT 549

Db 548 TATAAATAATATTAAWATTTGATTTTATTATTATTTTAAAAATTTTATTATATWA 489

Db 488 TTTATTAAATTTTANTATATATTTTTTATATCTTTCMAAAATATTTMTCCCYTTTWTM 429

Db 428 ATGTTTATTTTTHCTCTNCNTTTNTTTNTT 396

RESULTS	12		
CNS00EPO			
LOCUS	CNS00EPO	1101 bp	DNA
			GSS
			04-JUN-1999

BACR23P01 of RPCI-98 library from *Drosophila melanogaster* (fruit fly), genomic survey sequence.
 AL069493
 ACCESSION

KEYWORDS	fruit fly.
SOURCE	GSS.
ORGANISM	Drosophila melanogaster

REFERENCE

1 (bases 1 to 1101)

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

pterlygota; Neoptera; Endopterygota; Diptera; Brachycera;

TITLE
JOURNAL
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2002, 03:52:00 ; Search time 2120.21 Seconds
(without alignments)
60.819 Million cell updates/sec

Title: US-09-700-187-1

Perfect score: 12

Sequence: 1 ggattttacagt 12

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1:	em_estfun:*
2:	em_esthum:*
3:	em_estin:*
4:	em_estom:*
5:	em_estpl:*
6:	em_estba:*
7:	em_estrov:*
8:	em_estov:*
9:	em_hic:*
10:	gb_est1:*
11:	gb_est2:*
12:	gb_hic:*
13:	gb_gss:*
14:	em_gss_fun:*
15:	em_gss_hum:*
16:	em_gss_inv:*
17:	em_gss_pln:*
18:	em_gss_pro:*
19:	em_gss_rod:*
20:	em_gss_vrt:*
21:	em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	12	100.0	88	13	AZ925237 4910.ez32
2	12	100.0	118	10	AW217731 EST296445
3	12	100.0	126	13	AZ289196 RPCI-23-5
4	12	100.0	130	10	A1610647 tp20d02.x
5	12	100.0	147	10	BE145737 IL5-HT020
6	12	100.0	147	11	D25785 HUMGS04153
7	12	100.0	150	10	AA579315 nf36e06.s
8	12	100.0	152	10	AU180504 AU180504
9	12	100.0	155	13	AQ009485 CIT-HSP-2
10	12	100.0	165	10	AV333933 AV333933
11	12	100.0	165	11	BF881282 QV1-ET018
12	12	100.0	171	11	BF909645 PM3-UT005

c 13	12	100.0	172	10	AI029683
c 14	12	100.0	173	13	AZ818519
c 15	12	100.0	177	13	AQ985182
c 16	12	100.0	180	11	BF235518
c 17	12	100.0	184	10	AV054690
c 18	12	100.0	185	10	AV071084
c 19	12	100.0	185	10	BF357868
c 20	12	100.0	185	11	BG942699 ax28e01.x
c 21	12	100.0	186	10	AI345298
c 22	12	100.0	189	11	F29452
c 23	12	100.0	190	10	AV288772
c 24	12	100.0	193	10	AV285178
c 25	12	100.0	194	10	AI136924
c 26	12	100.0	202	11	D60959
c 27	12	100.0	207	10	AI572468
c 28	12	100.0	210	10	AV336005
c 29	12	100.0	216	13	B25087
c 30	12	100.0	217	10	AI563911
c 31	12	100.0	220	13	AQ278868
c 32	12	100.0	223	10	BE236288
c 33	12	100.0	223	10	BE236293
c 34	12	100.0	224	10	AV252038
c 35	12	100.0	224	10	BB512144
c 36	12	100.0	225	10	BB256670
c 37	12	100.0	226	11	BI220740
c 38	12	100.0	227	10	AA129692
c 39	12	100.0	227	11	BI293223
c 40	12	100.0	228	10	BB171532
c 41	12	100.0	228	10	BB310621
c 42	12	100.0	230	10	BB152264
c 43	12	100.0	231	10	AV368688
c 44	12	100.0	231	10	BB503511
c 45	12	100.0	231	10	BB510596

ALIGNMENTS

RESULT 1
AZ925237/c
LOCUS 88 bp DNA
DEFINITION 4910.ez32kl8.sl Saccharomyces paradoxus N17 Saccharomyces paradoxus genomic clone 4910.ez32kl8.sl, DNA sequence.
ACCESSION AZ925237
VERSION AZ925237.1 GI:13496136
KEYWORDS GSS.
SOURCE Saccharomyces paradoxus.
ORGANISM Saccharomyces paradoxus.
REFERENCE 1 (bases 1 to 88)
AUTHORS Cliften,P.F., Hillier,L.W., Fulton,L., Graves,T., Miner,T., Gish,W.R., Waterston,R.H. and Johnston,M.
TITLE Surveying Saccharomyces genomes to identify functional elements by comparative DNA sequence analysis
JOURNAL Unpublished (2001)
COMMENT Contact: Johnston M
Department of Genetics
Washington University Medical School
Box 8232, 4566 Scott Ave., St. Louis, MO 63110, USA
Tel: 314 362 2735
Fax: 314 362 7855
Email: mjgenetics.wustl.edu
Class: random plasmid subclone.
Location/Qualifiers
1..88
/organism="Saccharomyces paradoxus"
/strain="N17"
/db_xref="taxon:27291"
/clone="4910.ez32kl8.sl"
/clone_lib="Saccharomyces paradoxus N17"
/note="Random genomic sequence"
BASE COUNT 26 a 12 c 8 g 42 t

THIS PAGE BLANK (USPTO)

ORIGIN

Query Match 100.0%; Score 12; DB 13; Length 88;
 Best Local Similarity 100.0%; Pred. No. 7.3e+03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggattttacagt 12
 |||||
 Db 40 GGATTTTACAGT 29

RESULT 2

AW217731/c
 LOCUS
 DEFINITION EST296445 tomato flower buds 8 mm to pre-anthesis, Cornell University Lycopersicon esculentum cDNA clone cTOG6E5, mRNA

ACCESSION
 VERSION
 KEYWORDS
 SOURCE

ORGANISM

Lycopersicon esculentum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;

REFERENCE

1 (bases 1 to 118)
 van der Hoeven, R.S., Bezzerides, J.L., Matern, A.L., Holt, I.E., Liang, F., Upton, J., Hansen, T., Craven, M.B., Bowman, C.L., Ahn, S., Ronning, C.M., Fraser, C.M., Martin, G.B., Giovannoni, J.J., and Tanksley, S.D.
 Generation of ESTs from tomato flower tissue
 Unpublished (1999)

- TITLE

JOURNAL

COMMENT

FEATURES

source

Location/Qualifiers

1..118
 /organism="Lycopersicon esculentum"
 /cultivar="TA496"
 /db_xref="taxon:4081"
 /clone="cTOG6E5"
 /clone_lib="tomato flower buds 8 mm to pre-anthesis,
 Cornell University"
 /tissue_type="flower"
 /dev_stage="buds 8mm-to-preanthesis"
 /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
 XhoI; supplier: Tanksley; Flower buds and flowers were
 taken from greenhouse plants (4-8 wks old, TA496). They
 were immediately frozen in liquid nitrogen and then
 size-separated while remaining frozen."

BASE COUNT

ORIGIN

49 a 21 c 16 g 32 t

Query Match

Best Local Similarity 100.0%; Score 12; DB 10; Length 118;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggattttacagt 12
 |||||
 Db 47 GGATTTTACAGT 36

RESULT 3

AZ289196
 LOCUS
 DEFINITION RPCI-23-59B23.TJB RPCI-23 Mus musculus genomic clone RPCI-23-59B23,
 DNA sequence.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AZ289196.1 GI:9530982

GSS.

house mouse.

Mus musculus

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 126)

Zhao, S., Nieman, W., Feldblyum, T., Malek, J., Shatsman, S., Aknret,

B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.

and Fraser, C.M.

Mouse BAC End Sequences from Library RPCI-23

Unpublished (1999)

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)

or from Resea ch Genetics (info@resgen.com). BAC end page:

http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html

Plate: 59 row: B column: 23

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

1..126

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-23-59B23"

/clone_lib="RPCI-23"

/sex="Female"

/lab_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:

EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or

brain genomic DNA was isolated and partially digested

with a combination of EcoRI and EcoRI Methylase. Size

selected DNA was cloned into the pBACe3.6 vector at the

EcoRI sites. The ligation products were transformed into

DH10B electrocompetent cells (BRL Life Technologies).

38 a 20 c 30 g 38 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 12; Conservative

Mismatches 0; Indels

0; Gaps 0;

Qy 1 ggattttacagt 12

|||||

Db 93 GGATTTTACAGT 104

RESULT 4

AZ289196/c

LOCUS

DEFINITION

tp20d02.x1 NCI_CGAP_Gas4 Homo sapiens cDNA clone IMAGE:2188323 3'

similar to SW:ULC6_HCMVA P16836 HYPOPHRETHICAL PROTEIN UL126.

; contains element LTR5 repetitive element ; mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 130)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

THIS PAGE BLANK (USPTO)

JOURNAL COMMENT
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: crapsb-r@mail.nih.gov
 Tissue Procurement: Christopher Mokaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 cDNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality
 Seq primer: -40up from Gibco
 High quality sequence stop: 1.
FEATURES
 source
 1. .130
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="IMAGE:2188323"
 /clone_lib="NCL_CGAP_Gas4"
 /tissue_type="poorly differentiated adenocarcinoma with signet ring cell features"
 /lab_host="DH10B"
 /note="Organ: stomach; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt; Average insert size 1.69 kb. Life Technologies catalog #: 115497011"

BASE COUNT 36 a 21 c 28 g 45 t
 ORIGIN

Query Match 100.0%; Score 12; DB 10; Length 130;
 Best Local Similarity 100.0%; Pred. No. 7.1e+03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 ggattttacagt 12
 |||||
 Db 115 GGATTTTACAGT 104

RESULT 5
 BE145737/c
 LOCUS 147 bp mRNA EST 21-JUN-2000
DEFINITION
 IL5-HT0207-231099-006-A01 HT0207 Homo sapiens cDNA, mRNA sequence.
ACCESSION
 BE145737
VERSION
 BE145737.1 GI:8608461
KEYWORDS
 EST.
SOURCE
 human.
ORGANISM
 Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 147)
 Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zado,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL MEDLINE COMMENT
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-IL5-HT0207-231
 099-006-A01&t3=1999-10-23&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 56
 High quality sequence stop: 147.
 Location/Qualifiers

FEATURES
 source
 1. .147
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="HT0207"
 /dev_stage="Adult"
 /note="Organ: head neck; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
 BASE COUNT 31 a 35 c 12 g 69 t
 ORIGIN

Query Match 100.0%; Score 12; DB 10; Length 147;
 Best Local Similarity 100.0%; Pred. No. 7e+03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 ggattttacagt 12
 |||||
 Db 132 GGATTTTACAGT 121

RESULT 6
 D25785/c
 LOCUS 147 bp mRNA EST 30-NOV-1995
DEFINITION
 HUM504153 Human colon mucosa Homo sapiens cDNA clone cml984 3', mRNA sequence.
ACCESSION
 D25785
VERSION
 D25785.1 GI:500469
KEYWORDS
 EST.
SOURCE
 human.
ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 147)
 Okubo,K., Yoshii,J., Yokouchi,H., Kameyama,M. and Matsubara,K.
TITLE
 Global analysis of gene expression in colon mucosa: a large scale random cDNA sequencing analysis
 Unpublished (1994)
JOURNAL COMMENT
 Contact: Okubo,K., Itoh,K., Yoshii,J., Yokouchi,H. and Matsubara,K.
 Institute for Molecular and Cellular Biology
 Osaka University
 3-1 Yamada-oka,Suita,Osaka 565,Japan.
 Location/Qualifiers

FEATURES
 source
 1. .147
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="cml984"
 /clone_lib="Human colon mucosa"
 /note="Adult male, tissue_type = colon mucosa"
 BASE COUNT 63 a 24 c 34 g 26 t
 ORIGIN

Query Match 100.0%; Score 12; DB 11; Length 147;
 Best Local Similarity 100.0%; Pred. No. 7e+03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 ggattttacagt 12
 |||||
 Db 57 GGATTTTACAGT 46

THIS PAGE BLANK (USPTO)

```

ORGANISM
Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.

REFERENCE
1 (bases 1 to 152)

AUTHORS
Sanaka,E., Hori,H., Naruse,K., Mitani,H. and Tanaka,M.

TITLE
Medaka EST analysis

JOURNAL
Unpublished (2001)

COMMENT
Contact: Emi Sanaka
Department of Biological Sciences
Graduate School of Science, Nagoya University
Furo-cho, Chikusa-ku, Nagoya 464-8602, Japan
Tel: 81-52-789-2973
Fax: 81-52-789-2974
Email: sanaka@bio.nagoya-u.ac.jp

FEATURES
            Location/Qualifiers
                1..152
                    /organism="Oryzias latipes"
                    /strain="wild type"
                    /db_xref="taxon:8090"
                    /clone="NGY10.02c"
                    /clone_lib="Medaka eye cDNA library (SNK01)"
                    /tissue_type="eye"
                    /dev_stage="adult"
                    /note="Wild samples from Okayama Pref.(Southern part of
                        Japan)"

BASE COUNT      56 a   23 c   30 g   43 t
ORIGIN

Query Match      100.0%; Score 12; DB 10; Length 152;
Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy   +1 ggattttacagt 12
      |||||
Db    81 GGATTTTACAGT 92

RESULT 9
AQ009485
LOCUS
DEFINITION
CIT-HSP-2283N16.TRB CIT-HSP Homo sapiens genomic clone 2283N16, UNA
sequence.
ACCESSION      AQ009485
VERSION        AQ009485.1 GI:3128870
KEYWORDS
SOURCE        GSS.
SOURCE        human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 152)
Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
Golden,K., Berry,K., Granger,D., Suh,E., Wibie,C., Shizuwa,H.,
Simon,M. and Venter,J.C.
Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
Unpublished (1998)
Other_GSSs: CIT-HSP-2283N16.TFB
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: M13 Reverse
Class: BAC ends.

```

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV1&t2=QV1-ET0183-021200-529-d05_1&t3=2000-12-02&t4=1)
 Seq primer: puc 18 forward
 High quality sequence stop: 165.

FEATURES

Location/Qualifiers
 1..165
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="ET0183"
 /dev_stage="Adult"

/note="Organ: lung_tumor; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 34 a 46 c 35 g 50 t
 ORIGIN

Query Match 100.0%; Score 12; DB 11; Length 165;
 -Best Local Similarity 100.0%; Pred. No. 7e+03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gqattttacagt 12
 |||||
 Db 96 GGATTTACAGT 107

RESULT 12

LOCUS BF909645 171 bp mRNA EST 18-JAN-2001
 DEFINITION PM3-UT0058-181000-007-g01 UT0058 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BF909645
 VERSION BF909645.1 GI:12301103
 KEYWORDS EST.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS 1 (bases 1 to 171)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

TITLE

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL

PROC. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE

20202863

COMMENT

Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM3&t2=PM3-UT0058-181000-007-g01&t3=2000-10-18&t4=1>)

Seq primer: puc 18 forward

High quality sequence start: 17

High quality sequence stop: 113.

FEATURES

Location/Qualifiers
 1..171

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="UT0058"
 /dev_stage="Adult"
 /note="Organ: uterus_tumor; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 34 a 42 c 49 g 46 t
 ORIGIN

Query Match 100.0%; Score 12; DB 11; Length 171;
 -Best Local Similarity 100.0%; Pred. No. 7e+03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ugattttacagt 12
 |||||
 Db 117 GGATTTACAGT 128

RESULT 13

LOCUS AI029683 172 bp mRNA EST 04-JUL-1999
 DEFINITION UI-R-CO-ly-b-01-0-UI.s1 UI-R-C0 Rattus norvegicus cDNA, clone
 UI-R-CO-ly-b-01-0-UI 3', mRNA sequence.

ACCESSION AI029683
 VERSION AI029683.1 GI:4300245
 KEYWORDS EST.
 SOURCE Norway rat.

ORGANISM

Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE

1 (bases 1 to 172)
 Bonaldo, M.F., Lennon, G. and Soares, M.B.
 Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL

Genome Res. 6 (9), 791-806 (1996)

MEDLINE

9704477

COMMENT

On Jun 22, 1998 this sequence version replaced gi:3247509.
 Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized adult Liver library, cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics. This clone is also available through the I.M.A.G.E. Consortium at LNL (info@image.llnl.gov). IMAGE ID=1782730

Seq primer: M13 Forward

POLYA-No.

FEATURES

Location/Qualifiers
 1..172

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="UI-R-CO-ly-b-01-0-UI"

/clone_lib="UI-R-CO"

/dev_stage="adult"

/lab_host="DH10B (Life Technologies)"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CO

library is a subtracted library derived from the UI-R-A1

and UI-R-E1 libraries. The UI-R-A1 library consisted of a

mixture of individually tagged normalized libraries

THIS PAGE BLANK (USPTO)

constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, and muscle. The UI-R-EI library consisted of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-CO) was constructed as follows: PCR amplified cDNA inserts from a pool of UI-R-Al and UI-R-EI clones from which 3' ESTs had been derived was used as a driver in a hybridization with the pooled UI-R-Al and UI-R-EI library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-CO library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)*

BASE COUNT 32 a 47 c 32 g 61 t
ORIGIN

Query Match 100.0%; Score 12; DB 10; Length 172;
Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
|||||
Db 29 GGATTTTACAGT 18

RESULT 14

AZ818519 173 bp DNA GSS 20-FEB-2001
LOCUS
DEFINITION 2M0088A10R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC2M0088A10 R, DNA sequence.

ACCESSION AZ818519
VERSION AZ818519.1 GI:12988427
KEYWORDS
SOURCE

house mouse.
ORGANISM

REFERENCE
AUTHORS
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A., and Wright, D., Weiss, R.
TITLE
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL
COMMENT
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: dduwn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0088 row: A column: 10
Seq primer: CACACAGGAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 173.

FEATURES

source
1. .173
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0088A10"
/clone_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/note="vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 41 a 41 c 46 g 45 t
ORIGIN

Query Match 100.0%; Score 12; DB 13; Length 173;
Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
|||||
Db 134 GGATTTTACAGT 145

RESULT 15

AQ985182 177 bp DNA GSS 30-JAN-2000
LOCUS
DEFINITION RPCI-23-307M5.TJ RPCI-23 Mus musculus genomic clone RPCI-23-307M5, DNA sequence.

ACCESSION AQ985182
VERSION AQ985182.1 GI:6818387
KEYWORDS
SOURCE

house mouse.
ORGANISM

REFERENCE
AUTHORS
Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P., and Fraser, C.M.
TITLE
Mouse BAC End Sequences from Library RPCI-23

JOURNAL
COMMENT
Unpublished (1999)
Other GSSs: RPCI-23-307M5.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
Plate: 307 row: M column: 5
Seq primer: SP6
Class: BAC ends.

FEATURES

source
1. .177
/organism="Mus musculus"
/strain="C57BL/6J"

THIS PAGE BLANK (USPTO)

```
/db_xref="taxon:10090"  
/clone="RPCI-23-307M5"  
/clone_lib="RPCI-23"  
/sex="Female"  
/lab_host="DH10B"  
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:  
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or  
brain genomic DNA was isolated and partially digested  
with a combination of EcoRI and EcoRI Methylase. Size  
selected DNA was cloned into the pBACe3.6 vector at the  
EcoRI sites. The ligation products were transformed into  
DH10B electrocompetent cells (BRL Life Technologies)."
```

```
BASE COUNT      56 a      48 c      20 g      52 t      1 others  
ORIGIN  
  
Query Match      100.0%; Score 12; DB 13; Length 177;  
Best Local Similarity 100.0%; Pred. No. 7e+03;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY      1 ggattttacagt 12  
        |||||  
Db      139 GGATTTTACAGT 128
```

Search completed: April 3, 2002, 04:39:12
Job time: 2832 sec

THIS PAGE BLANK (USPTO)